

# Universität Rostock



Traditio et Innovatio

---

## Nichtparametrische Schätzer und Anpassungstests für Semi-Markov-Modelle

---

### Dissertation

zur

Erlangung des akademischen Grades  
doctor rerum politicarum (Dr. rer. pol.)

der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Universität Rostock

vorgelegt von

Lucas Raul Radloff,  
geb. am 05.04.1992 in Eutin  
aus Rostock

Rostock, 2021

**Gutachter**

- Prof. Dr. Rafael Weißbach  
Lehrstuhl für Statistik und Ökonometrie  
Institut für Volkswirtschaftslehre  
Universität Rostock
- Prof. Dr. Dominik Wied  
Institut für Statistik und Ökonometrie  
Universität zu Köln

**Jahr der Einreichung:** 2021

**Jahr des Kolloquiums:** 2021

# Danksagung

Mehr als drei Jahre sind eine lange Zeit und ich möchte mich bei all jenen Menschen bedanken, die mich in diesem Zeitraum unterstützt und damit zum Gelingen der vorliegenden Arbeit beigetragen haben.

Zuerst sei Professor Rafael Weißbach genannt, der diese Arbeit mit viel Einsatz betreut hat. Immer wieder hat er sich dabei viel Zeit für Gespräche und das Lesen meiner Arbeit genommen und war mit detailliertem wie konstruktivem Feedback eine große Hilfe.

Mit Dennis Müller, der ein Wegbegleiter seit Studienbeginn ist, habe ich unzählige fachliche wie nicht-fachliche Unterhaltungen geführt. Besonderer Dank gebührt ihm für seine umfangreichen Beiträge dazu, die Arbeit sprachlich und visuell auf das derzeitige Niveau zu heben.

Kristina Pieth war bei all den organisatorischen Fragen und Aufgaben, die im Laufe der Zeit aufgekommen sind, immer wieder hilfsbereit.

Meiner Freundin Evelyn, meinen Eltern Hannelore und Ralf sowie meinem Bruder Bastian möchte ich für den seelischen Rückhalt und all die aufmunternden Worte in jenen Phasen danken, in denen es weniger gut lief. Insbesondere denke ich aber an den schwierigen Abschnitt, den wir vor zwei Jahren zusammen durchleben mussten. Ihr habt diese Zeit so erträglich wie möglich gemacht. Ich vermisse dich, Papa.

Zu guter Letzt sei all jenen Freunden gedankt, die in den letzten drei Jahre für reichlich Ablenkung von allen fachlichen Belangen gesorgt haben: Benito und Bella, Oleg, Przemek, Steffan, René, John, Shane, Felix, Moritz – ich stoße auf euch an.

# Inhaltsverzeichnis

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Einleitung</b>   | <b>6</b>  |
| <b>2</b> | <b>Zwei Anpassungstests für Intensitätsmodelle</b>                  | <b>10</b> |
| 2.1      | Notation und Zusammenfassung der Resultate von Nielsen und Linton . | 10        |
| 2.1.1    | Modellrahmen . . . . .  | 10        |
| 2.1.2    | Schätzer . . . . .  | 12        |
| 2.1.3    | Sonstige Notation . . . . .   | 12        |
| 2.1.4    | Resultate . . . . .   | 13        |
| 2.2      | Konvergenz der endlichdimensionalen Verteilungen . . . . .          | 15        |
| 2.3      | Gleichmäßige Konvergenz des Varianzschätzers . . . . .              | 20        |
| 2.4      | Zwei Modellspezifikationstests . . . . .                            | 24        |
| <b>3</b> | <b>Semi-Markov-Modelle</b>  | <b>39</b> |
| 3.1      | Ein einfaches Modell mit drei Zuständen . . . . .                   | 42        |
| 3.1.1    | Daten . . . . .   | 45        |
| 3.1.2    | Zählprozess-Formulierung . . . . .                                  | 45        |
| 3.1.3    | Schätzer . . . . .  | 46        |
| 3.1.4    | Nachweise der Martingaleigenschaft . . . . .                        | 48        |
| 3.2      | Allgemeine Semi-Markov-Prozesse . . . . .                           | 56        |
| 3.2.1    | Daten und Schätzer . . . . .  | 61        |
| 3.2.2    | Nachweis der Martingaleigenschaft . . . . .                         | 62        |
| <b>4</b> | <b>Anwendung auf Daten</b>  | <b>71</b> |
| 4.1      | Die Daten . . . . .   | 71        |
| 4.2      | Das Modell . . . . .  | 72        |
| 4.3      | Ergebnisse der Intensitätsschätzung . . . . .                       | 73        |
| 4.4      | Ergebnisse der statistischen Tests . . . . .                        | 75        |
| 4.5      | Interpretation . . . . .  | 79        |
| <b>5</b> | <b>Diskussion</b>   | <b>81</b> |
| <b>A</b> | <b>R-Code</b>   | <b>88</b> |



# Kapitel 1

## Einleitung

Zeitstetige Mehrzustandsmodelle – also zeitstetige stochastische Prozesse mit endlichem Zustandsraum – haben seit geraumer Zeit Anwendung in diversen Disziplinen gefunden. Hougaard (1999) und Andersen u. Keiding (2002) etwa geben einen guten Überblick. Anwendungsbereiche umfassen die Epidemiologie (siehe etwa Commenges, 1999), Demographie, Biologie, Seismologie (z.B. Aspinall u. a., 2006), Zuverlässigkeitstechnik (etwa Bao u. a., 2019), Ökonomie und die Finanzwirtschaft (vergleiche Bluhm u. a., 2002, S. 197 ff.). Zwei Beispiele mögen der Illustration dienen: Ein klassisches Beispiel der Epidemiologie ist das (unidirektionale) Krankheit-Tod-Modell (in Anwendung z.B. in PROVA-Study-Group, 1991), illustriert in Abbildung 1.1. Ein Individuum ist zu Beginn gesund. Im Verlaufe der Zeit wird dieses Individuum irgendwann sterben und in den Zustand „tot“ übergehen. In der Zwischenzeit ist es möglich, aber nicht notwendig, dass das Individuum an einer gewissen Krankheit erkrankt. In dieser Version des Modells ist keine Rückkehr vom Zustand „krank“ in den Zustand „gesund“ möglich, es ist also zur Modellierung von chronischen Krankheiten geeignet.

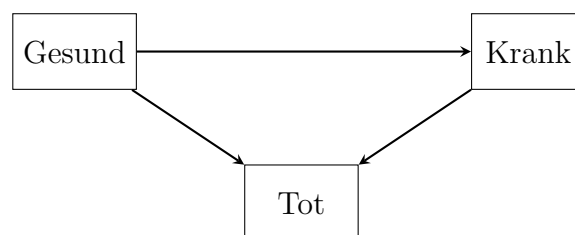


Abbildung 1.1: Strukturdiagramm des Krankheit-Tod-Modells. Pfeile zeigen an, welche Übergänge zwischen Zuständen möglich sind.

In der Finanzwirtschaft werden Übergänge zwischen den Ratingklassen eines Schuldners durch Mehrzustandsmodelle beschrieben (siehe etwa Weißbach u. Mollenhauer, 2011). Ein derartiges Rating beschreibt die Wahrscheinlichkeit, dass der Schuldner in Zukunft nicht in der Lage sein wird, seine Schulden zurückzuzahlen. Typischerweise reichen die Ratings vom bestmöglichen „AAA“ zum schlechtesten (außerhalb der Insolvenz) „C“

(etwa bei den Ratings von Standard & Poor's), siehe Abbildung 1.2.



Abbildung 1.2: Strukturdiagramm eines Kreditratingmodells. AAA entspricht dem besten Rating, der letzte Zustand D (wie Default), der der Insolvenz entspricht, ist absorbierend, d.h. er kann nicht mehr verlassen werden.

Eine bei der Modellierung mit Mehrzustandsmodellen häufig gemachte Annahme ist, dass es sich um einen Markovprozess handelt. Diese Annahme besagt, dass die zukünftige Verteilung eines Prozesses nur vom gegenwärtigen Zustand abhängt, nicht aber von anderen Informationen über die Vergangenheit. Im oben beschriebenen Krankheit-Tod-Modell würde dies etwa bedeuten, dass Sterbewahrscheinlichkeiten einer erkrankten Person nicht von der bisherigen *Dauer* der Erkrankung abhängen dürfen. Die mathematische Theorie vereinfacht sich durch die Markovannahme deutlich. Allerdings ist diese Annahme in vielen praktischen Situationen nicht tragbar. So finden etwa sowohl Pendlebury u. Rothwell (2009) als auch Corraini u. a. (2017) Hinweise darauf, dass die Zeit seit einem Schlaganfall ein Risikofaktor für einen Demenz-Ausbruch ist. Huo u. a. (2018) beschreiben Dauerabhängigkeit der Mortalität von Diabetes-Patienten und Lando u. Skødeberg (2002) sowie dos Reis u. a. (2020) äußern Zweifel an der Markovannahme für Rating-Modelle.

Eine Möglichkeit, mit dieser Problematik umzugehen, umfasst die Schätzung von Übergangswahrscheinlichkeiten gegeben nur den heutigen Zustand, wie es z.B. Titman (2015), Datta u. Satten (2001) oder Una-Alvarez u. Meira-Machado (2015) machen. Dieser Ansatz hat den Vorteil, dass wenige Annahmen über die Abhängigkeitsstruktur innerhalb des Prozesses gemacht werden. Allerdings werden bei den so geschätzten Wahrscheinlichkeiten potentiell relevante Informationen der Historie des Prozesses ignoriert. Daher kann es sinnvoll sein, stattdessen die zeitlichen Abhängigkeitsstrukturen explizit zu modellieren.

Eine Methode dafür ist es, zu Semi-Markovprozessen überzugehen. Siehe etwa Devolder u. a. (2015), Kapitel 4, für einen Überblick, oder Janssen u. Manca (2006) für weitere Details. Bei einem Semi-Markov-Prozess wird die Markovannahme dahingehend abgeschwächt, dass die zukünftige Verteilung des Prozesses neben dem heutigen Zustand auch von der bisherigen Verweildauer im aktuellen Zustand abhängen darf. Derartige Modelle werden in vielen Disziplinen angewandt, so z.B. von D'Amico u. a. (2017) im Finanzwesen, von Grabski (2015) in der Zuverlässigkeitstechnik oder von Biessy (2017) in einem aktuariellen Kontext.

Wir schlagen vor, einen nichtparametrischen kernbasierten Intensitätsschätzer unter Anwesenheit von Kovariaten, welcher von Nielsen u. Linton (1995) in einem allgemeinen Zählprozess-Kontext eingeführt wurde, zur Schätzung von Übergangsintensitäten

in zeitheterogenen Semi-Markov-Modellen zu nutzen. Die bisherige Verweildauer im aktuellen Zustand dient hier als derartige Kovariate. Um zu zeigen, dass die Theorie von Nielsen u. Linton (1995) anwendbar ist, betten wir Semi-Markov-Modelle in den Zählprozess-Kontext ein und weisen die Martingal-Eigenschaft eines damit zusammenhängenden Prozesses nach. Wir berücksichtigen außerdem Rechtszensierung, ein in Studien regelmäßig auftretendes Phänomen.

Darüber hinaus entwickeln wir zwei Anpassungstests basierend auf dem Intensitätsschätzer von Nielsen u. Linton (1995). Im Kontext der Regression existieren eine Reihe von Ansätzen, nichtparametrische kernbasierte Schätzer wie den Nadaraya-Watson-Schätzer als Grundlage von Anpassungstests zu nutzen, namentlich Gozalo (1993), Fan u. Li (1996), Lavergne u. Vuong (2000), Ait-Sahalia u. a. (2001) und Lavergne u. a. (2015). Im Kontext der Intensitätsschätzung unter Einfluss von Kovariaten ist uns aber keine allein auf einem Kernschätzer beruhende Methode bekannt. McKeague u. Utikal (1990a) beschreiben einen Test, bei dem die Teststatistik auf einer kumulierten Intensität beruht, der Anpassungstest von Vital u. Patil (2020) arbeitet in einem Setting ohne Kovariate. Wir entwickeln unseren Test im allgemeinen Zählprozess-Framework und adaptieren dazu einen Ansatz, der von Gozalo (1993) für Kernregressionsschätzer genutzt wurde. Im Kontext von Semi-Markov-Prozessen kann unser Test dazu verwendet werden, die Abhängigkeit einer Übergangsintensität von der Verweildauer im aktuellen Zustand zu überprüfen. Da ein Semi-Markov-Prozess genau dann markovsch ist, wenn (für alle möglichen Übergänge zwischen unterschiedlichen Zuständen) keine derartige Abhängigkeit vorliegt, kann so auch auf die Markov-Eigenschaft getestet werden.

Wir schließen die Arbeit ab, indem wir die Methoden an einem Datensatz vorführen. Wir verwenden Daten zu Demenzdiagnosen der deutschen gesetzlichen Krankenkasse AOK, siehe etwa auch Weißbach u. a. (2021). Von Interesse ist hier, inwiefern die Sterblichkeit von Personen mit Demenz von der bisherigen Dauer der Demenz abhängt.

Das Vorgehen strukturiert sich wie folgt: In Kapitel 2 entwickeln wir Anpassungstests für Intensitäten in einem allgemeinen Zählprozess-Rahmen. Dazu fassen wir zunächst die Ergebnisse von Nielsen u. Linton (1995) zusammen und verallgemeinern die Resultate um zwei Aussagen: die multivariate Normalität des Schätzers, welche wir mit Hilfe des Satzes von Cramér-Wold erhalten, und die gleichmäßige Konvergenz des Varianzschätzers. Diese Aussagen nutzen wir, um die Konsistenz zweier Anpassungstests nachzuweisen, die wir in Anlehnung an Gozalo (1993) definieren. In Kapitel 3 zeigen wir, dass sich die Schätzung von Übergangsintensitäten in Semi-Markov-Prozessen in den allgemeineren Zählprozess-Rahmen von Kapitel 2 eingliedern lässt. Dies liefert die Konsistenz des Intensitätsschätzers von Nielsen u. Linton (1995) sowie die Validität unserer Tests für Semi-Markov-Prozesse. Wir liefern den Nachweis zunächst für ein vereinfachtes progressives Modell mit drei Zuständen, jeweils ohne und mit Rechts-



zensierung, und im Anschluss für allgemeine zeitheterogene Semi-Markov-Prozesse. In Kapitel 4 folgt die Anwendung der Methoden auf Daten der AOK. Wir beenden die Arbeit in Kapitel 5 mit einer Diskussion der Resultate. Im Appendix schließlich ist der R-Code zu finden, welcher in der Anwendung für die Berechnung der Schätzer und Testergebnisse benutzt wurde.

# Kapitel 2

## Zwei Anpassungstests für Intensitätsmodelle

In diesem Kapitel werden wir – aufbauend auf einem nichtparametrischen, kernbasierten Schätzer für Übergangsintensitäten – zwei Anpassungstests konstruieren. Dazu werden wir zunächst die Resultate von Nielsen u. Linton (1995) (im Folgenden NL genannt), in deren Arbeit dieser Schätzer eingeführt wurde, zusammenfassen. Im Anschluss werden wir zwei zusätzliche Ergebnisse zu diesem Schätzer präsentieren: die multivariate asymptotische Normalität und die gleichmäßige Konsistenz der Varianzschätzers. Diese werden wir schließlich benötigen, um – angelehnt an Gozalo (1993), der einen ähnlichen Ansatz für Kern-Regressions-Schätzer vorgestellt hat – die Validität unserer Tests nachzuweisen.

### 2.1 Notation und Zusammenfassung der Resultate von Nielsen und Linton

Wir beginnen damit, das Framework und die Ergebnisse von NL zu erläutern, um innerhalb dieses Rahmens weiterarbeiten zu können.

#### 2.1.1 Modellrahmen

Es wird angenommen, dass  $n$  u.i.v. Beobachtungen  $(N_1, Z_1^*, Y_1), \dots, (N_n, Z_n^*, Y_n)$  vorliegen, wobei

- $N_i, i \in \mathbb{N}$ , Zählprozesse derart sind, dass  $N^{(n)} = (N_1, \dots, N_n)$  ein  $n$ -dimensionaler Zählprozess bezüglich einer rechtsstetigen, vollständigen Filtration  $\mathfrak{F}_t^{(n)} = \mathfrak{F}_t$  ist. Wie auch alle weiteren Prozesse ist  $N^{(n)}$  auf dem Einheitsintervall  $[0, 1]$  definiert.  $N_i$  zählt die Anzahl der beobachteten (Ausfall-)Ereignisse des  $i$ -ten Individuums.

- $Y_i, i \in \mathbb{N}$ , vorhersagbare Prozesse mit Werten in  $\{0, 1\}$  sind, die anzeigen, ob das  $i$ -te Individuum unter Risiko steht.
- $Z_i = (Z_{i1}, \dots, Z_{id}), i \in \mathbb{N}$ ,  $d$ -dimensionale, vorhersagbare CADLAG-Prozesse sind, welche als Kovariate bzw. Marker fungieren. Die tatsächlich beobachteten Variablen  $Z_i^*$  sind definiert durch

$$Z_i^*(s) := \begin{cases} Z_i(s), & \text{falls } Y_i(s) = 1 \\ -\infty, & \text{falls } Y_i(s) = 0 \end{cases},$$

d.h. wir beobachten die Kovariaten nur, solange das Individuum unter Risiko steht. Es wird weiterhin angenommen, dass  $Z_i(s)$  gegeben  $\{Y(s) = 1\}$  eine Lebesgue-stetige Verteilung mit dem Träger  $[0, 1]^d$  besitzt.

Die Wahl von  $[0, 1]$  als betrachtetem Zeitintervall bzw. Wertebereich der Kovariaten ist eine vereinfachende Konvention. Die Resultate gelten für beliebige kompakte Intervalle. NL nennen als Filtration  $\mathfrak{F}_t = \sigma(N_i(s), Z_i(s), Y_i(s); i = 1, \dots, n, s \leq t)$ . Diese spezifische Wahl ist allerdings nicht entscheidend. Vielmehr ist nur relevant, dass die Annahmen an die Filtration selbst sowie die Annahmen an die Prozesse, welche an die Filtration geknüpft sind, erfüllt werden. Diese Annahmen umfassen die Rechtsstetigkeit der Filtration, die Vorhersagbarkeit von  $Y_i$  und  $Z_i$  sowie die Martingaleigenschaft der später definierten Prozesse  $M_i$ . Modelliert wird nun der Intensitätsprozess  $\lambda^{(n)} = (\lambda_1, \dots, \lambda_n)$  von  $N^{(n)}$  in Abhängigkeit der Kovariaten:

$$\lambda_i(t) = \alpha(Z_i(t), t) \cdot Y_i(t)$$

Hierbei wird keine Annahme über die Form der Funktion  $\alpha : [0, 1]^{d+1} \rightarrow [0, \infty)$  gemacht. Lediglich Glattheitsvoraussetzungen werden später gestellt.

**Bemerkung:** Oben wird die Annahme formuliert, dass die Filtration  $\mathfrak{F}_t$  rechtsstetig und vollständig ist, d.h. den sogenannten *conditions habituelles* (siehe etwa Andersen u. a., 1993, S. 60) genügt. Die Filtration kann stets vervollständigt werden, sodass Vollständigkeit o.B.d.A. angenommen werden kann. Rechtsstetigkeit ist aber unter den hier formulierten Bedingungen für  $\mathfrak{F}_t = \sigma(N_i(s), Z_i(s), Y_i(s); i = 1, \dots, n, s \leq t)$  nicht garantiert und muss im Einzelfall nachgeprüft werden. Die Theorie beruht allerdings auch nicht auf dieser spezifischen Filtration, sondern nutzt nur die genannten Eigenschaften, da diese Voraussetzung für einige gängige Theoreme sind, die benötigt werden. Daher kann für einen Anwendungsfall auch auf eine abgeänderte Filtration zugegriffen werden, bezüglich derer alle notwendigen Eigenschaften erfüllt sind.

Es sei ein Beispiel genannt, um zu illustrieren, dass die Rechtsstetigkeit verletzt sein kann. Dazu sei  $A \in \mathfrak{F}$  stochastisch unabhängig von den Prozessen  $N$  und  $Y$  sowie

$t_0 \in [0, 1]$  fix. Es sei der Prozess  $Z$  definiert durch

$$Z(t) := \mathbb{1}_A \mathbb{1}_{(t_0, \infty)}(t).$$

$Z$  besitzt linksstetige Pfade und ist daher bzgl.  $\mathfrak{F}_t = \sigma(N(s), Z(s), Y(s), s \leq t)$  vorhersagbar. Für das Ereignis  $A$  gilt  $A \in \mathfrak{F}_{t_0+h}$  für alle  $h > 0$  und somit ebenfalls  $A \in \bigcap_{h>0} \mathfrak{F}_{t_0+h} = \mathfrak{F}_{t_0+}$ . Da allerdings  $A \notin \mathfrak{F}_{t_0}$ , ist die Rechtsstetigkeit verletzt.

### 2.1.2 Schätzer

Für die Schätzung sei  $k$  eine eindimensionale Dichtefunktion und für die Bandbreite  $b > 0$  sei  $k_b(\cdot) = b^{-1}k(\cdot/b)$ . Weiterhin seien  $K(u) = \prod_{j=1}^{d+1} k(u_j)$  und  $K_b(u) = \prod_{j=1}^{d+1} k_b(u_j)$ . Identische Kerne und Bandbreiten für alle Kovariaten werden nur zum Zwecke der Übersichtlichkeit verwendet und sind für die Theorie nicht notwendig. Für  $z \in [0, 1]^d$  und  $t \in [0, 1]$  wird die Bezeichnung  $x = (z, t)$  genutzt. Analog sei  $X_i(s) := (Z_i(s), s)$ ,  $i = 1, \dots, n$ ,  $s \in [0, 1]$ , definiert.

Ziel ist die Schätzung der Funktion  $\alpha$  aus den Daten. NL schlagen hierfür folgenden Schätzer vor, welcher eine Verallgemeinerung eines Schätzers von Hjort (1994) für den Fall ohne Kovariaten darstellt:

$$\hat{\alpha}(z, t) := \hat{\alpha}(x) := \frac{\sum_{i=1}^n \int K_b(x - X_i(s)) dN_i(s)}{\sum_{i=1}^n \int K_b(x - X_i(s)) Y_i(s) ds} \quad (2.1)$$

Hier und in der gesamten übrigen Arbeit bezeichnen Integrale ohne notierte Grenzen Integrale über das Intervall  $[0, 1]$ .

NL weisen darauf hin, dass es sich hierbei im Sinne von Jones u. a. (1994) um einen „externen“ Schätzer handelt, was bedeutet, dass die Normierung außerhalb der Summe im Zähler erfolgt. Insofern ähnelt er in seiner Struktur dem Nadaraya-Watson-Schätzer für Regressionsfunktionen. Kernglättungen des Nelson-Aalen-Schätzers (Ramlau-Hansen, 1983) und Verallgemeinerungen dieser für die Einbeziehung von Kovariaten (siehe etwa McKeague u. Utikal, 1990b) wären hingegen „interne“ Schätzer.

### 2.1.3 Sonstige Notation

Es bezeichne  $f(z, t)$  die bedingte Dichte von  $Z_i(t)$  gegeben  $\{Y_i(t) = 1\}$  und  $y(t) = \mathbb{E}Y_i(t) = P(Y_i(t) = 1)$ . Weiterhin sei  $\varphi(x) = \varphi(z, t) = f(z, t)y(t)$ .

Die Prozesse  $M_i$ ,  $i = 1, \dots, n$ , werden definiert durch  $M_i(t) := N_i(t) - \int_0^t \lambda_i(s) ds$ . Gemäß den obigen Annahmen handelt es sich bei diesen um quadratintegrierbare lokale Martingale auf dem Intervall  $[0, 1]$ .

Außerdem definieren wir die Kernmomente  $\kappa_1 := \int_{-1}^1 v^2 k(v) dv$  und  $\kappa_2 := \int_{-1}^1 k(v)^2 dv$ .

Mit der Definition

$$\alpha^*(x) := \frac{\sum_{i=1}^n \int K_b(x - X_i(s)) \alpha(X_i(s)) Y_i(s) ds}{\sum_{i=1}^n \int K_b(x - X_i(s)) Y_i(s) ds} \quad (2.2)$$

lässt sich die Differenz  $(\hat{\alpha} - \alpha)(x)$  in zwei Summanden zerlegt,

$$(\hat{\alpha} - \alpha)(x) = (\hat{\alpha} - \alpha^*)(x) + (\alpha^* - \alpha)(x) = \frac{\mathcal{V}_x + \mathcal{B}_x}{\mathcal{C}_x}, \quad (2.3)$$

wobei

$$\begin{aligned} \mathcal{V}_x &= \frac{1}{n} \sum_{i=1}^n \int K_b(x - X_i(s)) dM_i(s) \\ \mathcal{B}_x &= \frac{1}{n} \sum_{i=1}^n \int K_b(x - X_i(s)) [\alpha(X_i(s)) - \alpha(x)] Y_i(s) ds \\ \mathcal{C}_x &= \frac{1}{n} \sum_{i=1}^n \int K_b(x - X_i(s)) Y_i(s) ds. \end{aligned}$$

NL bezeichnen  $(\hat{\alpha} - \alpha^*)(x)$  als variablen Term und  $(\alpha^* - \alpha)(x)$  als stabilen Term.

## 2.1.4 Resultate

Wir wollen drei Hilfsresultate sowie die beiden Hauptresultate von NL nennen, die wir im Folgenden noch verwenden.

### Hilfsresultate

Die drei Lemmata finden sich in NL als Proposition 1, Formel (10) sowie Lemma 1.

#### Lemma 2.1.1

Für alle  $n \in \mathbb{N}$  seien  $H_i^{(n)}, i = 1, \dots, n$ , vorhersagbare stochastische Prozesse auf dem Intervall  $[0, 1]$ . Für  $n \rightarrow \infty$  seien die folgenden Annahmen erfüllt:

(G1) Für  $\sigma^2 > 0$  gilt

$$\sum_{i=1}^n \int \left\{ H_i^{(n)} \right\}^2(s) d\langle M_i \rangle(s) \longrightarrow_P \sigma^2.$$

(G2) Für alle  $\varepsilon > 0$  gelte

$$\sum_{i=1}^n \int \left\{ H_i^{(n)} \right\}^2(s) \mathbb{1}_{\{|H_i^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \longrightarrow_P 0.$$

Dann folgt, dass

$$\sum_{i=1}^n \int H_i^{(n)}(s) dM_i(s) \Longrightarrow N(0, \sigma^2).$$

Dabei bezeichnet  $\rightarrow_P$   $P$ -stochastische Konvergenz und  $\implies$  schwache Konvergenz der Verteilungen der Zufallsvariablen auf der linken Seite gegen die Verteilung auf der rechten Seite. Die gleichen Bezeichnungen werden später auch für Zufallsvektoren genutzt.

**Lemma 2.1.2**

Es sei  $g_n(x)$  eine Folge reellwertiger Funktionen auf  $[0, 1]^{d+1}$  und  $G_{ni} = \int g_n(Z_i(s), s) Y_i(s) ds$ . Dann gilt

$$\text{Var} \left( \frac{1}{n} \sum_{i=1}^n G_{ni} \right) \leq \frac{1}{n} \int_{[0,1]^{d+1}} g_n(x)^2 \varphi(x) dx.$$

**Lemma 2.1.3**

Sei  $X(t)$  ein stochastischer Prozess mit  $t = (t_1, \dots, t_d) \in [0, 1]^d$ . Für  $t \in [0, 1]^d$  und  $v \in [0, 1]$ , sei  $t_{j,v} = (t_1, \dots, t_{j-1}, v, t_{j+1}, \dots, t_d)$ . Für ein  $C > 0$  gelten

(L1)  $X(t) \rightarrow_P 0$  für alle  $t \in [0, 1]^d$  und

(L2)  $\mathbb{E} [X(t) - X(t_{j,v})]^2 \leq C |t_j - v|^2$  für alle  $t \in [0, 1]^d, v \in [0, 1]$  und  $j \in \{1, \dots, d\}$ .

Dann folgt, dass

$$\sup_{t \in [0,1]^d} |X(t)| \rightarrow_P 0.$$

**Hauptresultate**

Die hier aufgeführten Sätze sind in NL als Theorem 1 bzw. 2 zu finden.

Für  $x \in (0, 1)^{d+1}$  sei eine Umgebung  $\mathcal{N} := [x - \varepsilon, x + \varepsilon] \subset (0, 1)^{d+1}$  definiert mit  $\varepsilon \in (0, 1)^{d+1}$ . Folgende Annahmen werden formuliert:

(D) Die Funktion  $\varphi$  ist strikt positiv auf  $\mathcal{N}$ .

(S) Die Funktion  $\alpha$  ist zweimal und die Funktion  $\varphi$  ist einmal stetig differenzierbar auf  $\mathcal{N}$ .

(K) Die Kernfunktion  $k$  besitzt den Träger  $[-1, 1]$ , ist symmetrisch um null und ist stetig.

(B) Für  $n \rightarrow \infty$  gelten  $nb^{d+1} \rightarrow \infty$  und  $b \rightarrow 0$ .

**Satz 2.1.4** (Punktweise Konvergenz)

Unter den Bedingungen (D), (S), (K) und (B) gelten die folgenden Aussagen:

(a)

$$n^{1/2} b^{(d+1)/2} \{\hat{\alpha}(x) - \alpha^*(x)\} \implies N \left[ 0, \kappa_2^{d+1} \frac{\alpha(x)}{\varphi(x)} \right]$$

(b)

$$b^{-2}\{\alpha^*(x) - \alpha(x)\} \longrightarrow_P \kappa_1 \sum_{j=1}^{d+1} \left\{ \frac{(\partial\alpha(x)/\partial x_j)(\partial\varphi(x)/\partial x_j)}{\varphi(x)} + \frac{\partial^2\alpha(x)/\partial x_j^2}{2} \right\}$$

(c)

$$\hat{\sigma}_x^2 = \mathcal{C}_x^{-2} n^{-1} b^{d+1} \sum_{i=1}^n \int K_b^2(x - X_i(s)) dN_i(s) \longrightarrow_P \sigma_x^2 = \kappa_2^{d+1} \frac{\alpha(x)}{\varphi(x)}$$

Nun sei  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_{d+1}$  ein  $(d+1)$ -dimensionales Teilintervall von  $[0, 1]^{d+1}$ , sodass jedes  $\mathcal{X}_i$  ein kompaktes Intervall ist. Es seien vier weitere Annahmen formuliert:

(D') Es gilt  $\inf_{x \in \mathcal{X}} \varphi(x) > 0$ .

(S') Es sind  $\alpha$  und  $\varphi$  stetig auf  $\mathcal{X}$ .

(K') Es gilt (K). Außerdem sei  $k$  Lipschitz-stetig, d.h. es existiert ein  $C > 0$ , sodass  $|k(u) - k(v)| \leq C|u - v|$  für alle  $u, v$ .

(B') Für  $n \rightarrow \infty$  gelten  $nb^{d+3} \rightarrow \infty$  und  $b \rightarrow 0$ .

**Satz 2.1.5** (Globale Konvergenz)

Unter den Annahmen (D'), (S'), (K') und (B') gilt

$$\sup_{x \in \mathcal{X}} |(\hat{\alpha} - \alpha)(x)| \longrightarrow_P 0.$$

## 2.2 Konvergenz der endlichdimensionalen Verteilungen

In Theorem 2.1.4 wurde die Normalität des Schätzers an einzelnen Punkten  $x \in (0, 1)^{d+1}$  gezeigt. Diese Aussage soll nun verallgemeinert werden auf endlich viele Punkte  $x_1, \dots, x_m \in (0, 1)^{d+1}$ . Wir stellen fest, dass asymptotisch Unabhängigkeit der Schätzungen an mehreren Punkten eintritt. Dies ist insofern ein erwartbares Resultat, als dass analoge Aussagen auch für Kerndichteschätzer und Kernregressionsschätzer gelten (siehe etwa Nadaraya, 1989, S. 88 bzw. S. 120). Zum Beweis dieser Aussage stellen wir zwei Lemmata voran.

**Lemma 2.2.1**

Für reellwertige Funktionen  $g$  gilt

$$\mathbb{E} \int g(Z_i(s), s) Y_i(s) ds = \int_{[0,1]^{d+1}} g(w) \varphi(w) dw.$$

**Beweis:**

$$\begin{aligned}
 \mathbb{E} \int g(Z_i(s), s) Y_i(s) ds &\stackrel{a.}{=} \int_{\Omega} \int g(Z_i(s), s) Y_i(s) ds dP \stackrel{b.}{=} \int \left( \int_{\Omega} g(Z_i(s), s) Y_i(s) dP \right) ds \\
 &\stackrel{c.}{=} \int \left( \int_{\{0,1\}} \int_{[0,1]^d} g(z, s) y \, dP_{Z(s)|Y(s)=y}(z) dP_Y(y) \right) ds \\
 &\stackrel{d.}{=} \int \left( \int_{[0,1]^d} g(z, s) f(z, s) y(s) dz \right) ds \stackrel{e.}{=} \int_{[0,1]^{d+1}} g(w) \varphi(w) dw.
 \end{aligned}$$

a. Definition des Erwartungswertes.

b. Satz von Fubini.

c. Maß-Transformationssatz.

d. Einsetzen von Dichten.

e. Verkürzen der Notation durch  $w = (z, s)$  und der Definition von  $\varphi$ . □

**Lemma 2.2.2**

Es sei  $m \in \mathbb{N}$ . Für alle  $n \in \mathbb{N}$  seien  $H_{i,j}^{(n)}, i = 1, \dots, n, j = 1, \dots, m$ , vorhersagbare stochastische Prozesse auf dem Intervall  $[0, 1]$ . Für  $n \rightarrow \infty$  seien die folgenden Annahmen erfüllt:

(G1') Für eine positiv definite, symmetrische  $m \times m$ -Matrix  $A = \{a_{jk}\}_{j,k=1,\dots,m}$  gilt

$$\sum_{i=1}^n \int H_{i,j}^{(n)}(s) H_{i,k}^{(n)}(s) d\langle M_i \rangle(s) \longrightarrow_P a_{jk}, \quad j, k = 1, \dots, m.$$

(G2') Für alle  $\varepsilon > 0$  gelte

$$\sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2(s) \mathbb{1}_{\{|H_{i,j}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \longrightarrow_P 0, \quad j = 1, \dots, m.$$

Dann folgt, dass

$$\left[ \sum_{i=1}^n \int H_{i,j}^{(n)}(s) dM_i(s) \right]_{j=1,\dots,m} \Longrightarrow N(\mathbf{0}, A).$$

**Bemerkung:** Lemma 2.2.2 ist eine  $m$ -dimensionale Verallgemeinerung des Lemmas 2.1.1. Er verhält sich zu diesem ähnlich wie Theorem I.2 in Andersen u. Gill (1982) zur klassischen prozesswertigen Variante des Grenzwertsatzes von Rebolledo (Rebolledo, 1980). Den Beweis von Lemma 2.2.2 schieben wir bis zum Ende des Beweises des eigentlichen Resultats in Satz 2.2.3 auf.



**Satz 2.2.3**

Für jeden der paarweise verschiedenen Punkte  $x_1, \dots, x_m \in (0, 1)^{d+1}$  seien die Annahmen (D) und (S) erfüllt. Außerdem gelten (K) und (B). Dann gilt

$$n^{1/2}b^{(d+1)/2}(\hat{\alpha} - \alpha^*)(\mathbf{x}) \implies N\left(\mathbf{0}, \text{diag}\left(\kappa_2^{d+1} \frac{\alpha(x_j)}{\varphi(x_j)}, j = 1, \dots, m\right)\right).$$

Für Konfidenzintervalle und Tests ist asymptotische Normalität nicht um  $\alpha^*$ , sondern um  $\alpha$  von Interesse. NL weisen bereits darauf hin, dass man diese bei „Unterglätten“ erhält; das heißt, indem man für wachsenden Stichprobenumfang die Bandbreite schneller klein werden lässt, als es im Sinne geringster quadratischer Abweichung optimal wäre. Um dies zu formalisieren formulieren wir

( $\tilde{B}$ ) Für  $n \rightarrow \infty$  gilt  $nb^{d+5} \rightarrow 0$ .

Diese zusätzliche Annahme garantiert nämlich anschaulich, dass

$$\frac{\text{Größe des stabilen Terms } (\alpha^*(x) - \alpha(x))}{\text{Breite des Konfidenzintervalls}} \longrightarrow 0.$$

Formalisiert wird der Gedanke durch folgendes Lemma.

**Korollar 2.2.4**

Es seien die Annahmen des Satzes 2.2.3 sowie ( $\tilde{B}$ ) erfüllt. Dann gilt sogar

$$n^{1/2}b^{(d+1)/2}(\hat{\alpha} - \alpha)(\mathbf{x}) \implies N\left(\mathbf{0}, \text{diag}\left(\kappa_2^{d+1} \frac{\alpha(x_j)}{\varphi(x_j)}, j = 1, \dots, m\right)\right).$$

**Beweis:** Wir stellen fest, dass

$$n^{1/2}b^{(d+1)/2}(\hat{\alpha} - \alpha)(\mathbf{x}) = n^{1/2}b^{(d+1)/2}(\hat{\alpha} - \alpha^*)(\mathbf{x}) + n^{1/2}b^{(d+1)/2}b^2b^{-2}(\alpha^* - \alpha)(\mathbf{x}).$$

Der erste Summand konvergiert nach Satz 2.2.3 gegen die Zielverteilung. Nach Satz 2.1.4 konvergiert  $b^{-2}(\alpha^* - \alpha)(\mathbf{x})$  gegen einen konstanten Vektor, da stochastische Konvergenz eines Vektors äquivalent zu stochastischen Konvergenz seiner Komponenten ist. Der Vorfaktor  $n^{1/2}b^{(d+1)/2}b^2 = (nb^{d+5})^{1/2}$  konvergiert gemäß ( $\tilde{B}$ ) gegen null.  $\square$

**Beweis Satz 2.2.3:** Lemma 2.2.2 soll angewendet werden mit

$$H_{i,j}^{(n)}(t) = n^{-1/2}b^{(d+1)/2}K_b(x_j - X_i(t)), \quad j = 1, \dots, m, i = 1, \dots, n.$$

Wir notieren außerdem, dass  $d\langle M_i \rangle(s) = \alpha(s, Z_i(s))Y_i(s)ds$  (siehe etwa Andersen u. a., 1993, S. 74, Formel (2.4.3)). Wir wollen (G1') nachweisen, indem wir zunächst die Konvergenz der Erwartungswerte zeigen und im Anschluss prüfen, dass die Varianzen

asymptotisch verschwinden. Es gilt

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{i=1}^n \int H_{i,j}^{(n)}(s) H_{i,k}^{(n)}(s) d\langle M_i \rangle(s) \right] \\
& \stackrel{a.}{=} b^{d+1} \int_{[0,1]^{d+1}} K_b(x_l - w) K_b(x_j - w) \alpha(w) \varphi(w) dw \\
& \stackrel{b.}{=} \int_{-x_j/b}^{(1-x_j)/b} K(q) K\left(\frac{x_j - x_l}{b} + q\right) \alpha(x_j + bq) \varphi(x_j + bq) dq \\
& \stackrel{c.}{\rightarrow} \begin{cases} \kappa_2^{d+1} \alpha(x_l) \varphi(x_l) & l = j \\ 0 & l \neq j \end{cases}.
\end{aligned}$$

- Wegen  $d\langle M_i \rangle(s) = \alpha(s, Z_i(s)) Y_i(s) ds$  liefert Lemma 2.2.1 die Umformung.
- Wir führen die Substitution  $q = (x_j - w)/b$  durch und nutzen die Symmetrie von  $K$  aus.
- Der Fall  $l = j$  ist in NL bereits abgedeckt. Wir stellen zunächst fest, dass für hinreichend große  $n$ , also kleine  $b$ , der Integrationsbereich den Träger von  $K$  enthält und wir somit über  $[-1, 1]^{d+1}$  integrieren. Für hinreichend kleine  $b$  ist  $K((x_l - x_j)/b + q) = 0$  für alle  $q \in [-1, 1]^{d+1}$ . Somit ist der gesamte Integrand und damit auch das Integral identisch null.

Für die Varianz erhalten wir

$$\begin{aligned}
& Var \left[ \sum_{i=1}^n \int H_{i,j}^{(n)}(s) H_{i,k}^{(n)}(s) d\langle M_i \rangle(s) \right] \\
& \stackrel{a.}{\leq} \frac{1}{n} b^{2d+2} \int_{[0,1]^{d+1}} K_b^2(x_l - w) K_b^2(x_j - w) \alpha^2(w) \varphi(w) dw \\
& \stackrel{b.}{=} \frac{1}{n} b^{2d+2-4d-4+d+1} \int_{-x_j/b}^{(1-x_j)/b} K^2(q) K^2\left(\frac{x_j - x_l}{b} + q\right) \alpha^2(x_j + bq) \varphi(x_j + bq) dq \\
& \stackrel{c.}{=} o(1).
\end{aligned}$$

- Lemma 2.1.2.
- Wir führen die Substitution  $q = (x_j - w)/b$  durch und nutzen die Symmetrie von  $K$  aus.
- Für hinreichend große  $n$  enthält der Integrationsbereich den Träger von  $K$  und wir integrieren somit über  $[0, 1]^{d+1}$ . Da alle Integranden auf diesem Bereich beschränkt sind, bleibt auch das Integral beschränkt. Der Vorfaktor ist  $o(1)$  gemäß Annahme (B). Für  $j \neq l$  ist das Integral sogar identisch null für hinreichend große  $n$  aus denselben Gründen, die für die Erläuterung des Erwartungswertes genutzt wurden.

(G2') ist bereits durch NL abgedeckt ist, ebenso wie  $\mathcal{C}_x \rightarrow_P \varphi(x)$  für alle  $x_j, j = 1, \dots, m$ . Somit ist Satz 2.2.3 gezeigt.  $\square$

Es bleibt, den Beweis von Lemma 2.2.2 nachzuholen:

**Beweis von Lemma 2.2.2:** Die Aussage folgt mit Cramér-Wold-Argumenten (siehe etwa Theorem 29.4 in Billingsley (2012)) aus Lemma 2.1.1. Es seien dafür  $c_1, \dots, c_m \in \mathbb{R}$ . Zunächst gilt

$$\sum_{j=1}^m c_j \sum_{i=1}^n \int H_{i,j}^{(n)}(s) dM_i(s) = \sum_{i=1}^n \int \left[ \sum_{j=1}^m c_j H_{i,j}^{(n)}(s) \right] dM_i(s).$$

Dessen Konvergenz in Verteilung für  $n \rightarrow \infty$  soll mit Hilfe von Lemma 2.1.1 nachgewiesen werden mit  $H_i^{(n)} = \sum_{j=1}^m H_{i,j}^{(n)}$ . Aus (G1') folgt, dass

$$\begin{aligned} \sum_{i=1}^n \int \left[ \sum_{j=1}^m c_j H_{i,j}^{(n)}(s) \right]^2 d\langle M_i \rangle(s) &= \sum_{i=1}^n \int \sum_{j=1}^m \sum_{k=1}^m c_j c_k H_{i,j}^{(n)}(s) H_{i,k}^{(n)}(s) d\langle M_i \rangle(s) \\ &= \sum_{j=1}^m \sum_{k=1}^m c_j c_k \sum_{i=1}^n \int H_{i,j}^{(n)}(s) H_{i,k}^{(n)}(s) d\langle M_i \rangle(s) \rightarrow_P \sum_{j=1}^m \sum_{k=1}^m c_j c_k a_{jk}. \end{aligned}$$

Damit ist (G1) gezeigt mit  $\sigma^2 = \sum_{j=1}^m \sum_{k=1}^m c_j c_k a_{jk}$ . Weiterhin erhalten wir

$$\begin{aligned} \sum_{i=1}^n \int \left[ \sum_{j=1}^m c_j H_{i,j}^{(n)}(s) \right]^2 \mathbb{1}_{\{|\sum_{l=1}^m c_l H_{i,l}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) &= \sum_{j=1}^m \sum_{k=1}^m c_j c_k \sum_{i=1}^n \int H_{i,j}^{(n)}(s) H_{i,k}^{(n)}(s) \mathbb{1}_{\{|\sum_{l=1}^m c_l H_{i,l}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \\ &\leq \sum_{j=1}^m \sum_{k=1}^m c_j c_k \sum_{i=1}^n \left[ \int \left\{ H_{i,j}^{(n)} \right\}^2(s) \mathbb{1}_{\{|\sum_{l=1}^m c_l H_{i,l}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \right]^{\frac{1}{2}} \cdot \\ &\quad \left[ \int \left\{ H_{i,k}^{(n)} \right\}^2(s) \mathbb{1}_{\{|\sum_{l=1}^m c_l H_{i,l}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \right]^{\frac{1}{2}} \\ &\leq \sum_{j=1}^m \sum_{k=1}^m c_j c_k \left[ \sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2(s) \mathbb{1}_{\{|\sum_{l=1}^m c_l H_{i,l}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \right]^{\frac{1}{2}} \cdot \\ &\quad \left[ \sum_{i=1}^n \int \left\{ H_{i,k}^{(n)} \right\}^2(s) \mathbb{1}_{\{|\sum_{l=1}^m c_l H_{i,l}^{(n)}(s)| > \varepsilon\}} d\langle M_i \rangle(s) \right]^{\frac{1}{2}}, \end{aligned}$$

wobei beide Abschätzungen Anwendungen der Cauchy-Schwarz-Ungleichung sind. Im ersten Fall wird diese auf die Integrale angewandt, was möglich ist, da  $\langle M_i \rangle(s)$  monoton nichtfallend ist. Im zweiten Fall findet die Anwendung auf die Summe über  $i$  statt.

Damit genügt es zu zeigen, dass

$$\sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2 (s) \mathbb{1}_{\left\{ \sum_{l=1}^m |c_l H_{i,l}^{(n)}(s)| > \varepsilon \right\}} d\langle M_i \rangle(s) \xrightarrow{P} 0, \quad j = 1, \dots, m,$$

um Bedingung (G2) nachzuweisen. Für diese Terme erhalten wir mit  $\tilde{\varepsilon} = \varepsilon / (m \cdot \max_l |c_l|)$

$$\begin{aligned} & \sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2 (s) \mathbb{1}_{\left\{ \sum_{l=1}^m |c_l H_{i,l}^{(n)}(s)| > \varepsilon \right\}} d\langle M_i \rangle(s) \\ & \stackrel{a.}{\leq} \sum_{i=1}^n \sum_{l=1}^m \int \left\{ H_{i,j}^{(n)} \right\}^2 (s) \mathbb{1}_{\{|H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}} d\langle M_i \rangle(s) \\ & \stackrel{b.}{=} \sum_{l=1}^m \left[ \sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2 (s) \mathbb{1}_{\{|H_{i,j}^{(n)}(s)| > \tilde{\varepsilon}, |H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}} d\langle M_i \rangle(s) \right. \\ & \quad \left. + \sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2 (s) \mathbb{1}_{\{|H_{i,j}^{(n)}(s)| \leq \tilde{\varepsilon}, |H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}} d\langle M_i \rangle(s) \right] \\ & \stackrel{c.}{\leq} \sum_{l=1}^m \left[ \sum_{i=1}^n \int \left\{ H_{i,j}^{(n)} \right\}^2 (s) \mathbb{1}_{\{|H_{i,j}^{(n)}(s)| > \tilde{\varepsilon}\}} d\langle M_i \rangle(s) \right. \\ & \quad \left. + \sum_{i=1}^n \int \left\{ H_{i,l}^{(n)} \right\}^2 (s) \mathbb{1}_{\{|H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}} d\langle M_i \rangle(s) \right] \xrightarrow{d.} 0. \end{aligned}$$

- a. Aus  $\sum_{l=1}^m |c_l H_{i,l}^{(n)}(s)| > \varepsilon$  folgt für mindestens ein  $l \in \{1, \dots, m\}$ , dass  $|H_{i,l}^{(n)}(s)| > \varepsilon / (m |c_l|)$ . Damit gilt dann auch insbesondere  $\mathbb{1}_{\left\{ \sum_{l=1}^m |c_l H_{i,l}^{(n)}(s)| > \varepsilon \right\}} \leq \sum_{l=1}^m \mathbb{1}_{\{|H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}}$ .
- b.  $\mathbb{1}_{\{|H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}} = \mathbb{1}_{\{|H_{i,j}^{(n)}(s)| > \tilde{\varepsilon}, |H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}} + \mathbb{1}_{\{|H_{i,j}^{(n)}(s)| \leq \tilde{\varepsilon}, |H_{i,l}^{(n)}(s)| > \tilde{\varepsilon}\}}$ .
- c. Auf dem Ereignis  $\left\{ |H_{i,j}^{(n)}(s)| \leq \tilde{\varepsilon}, |H_{i,l}^{(n)}(s)| > \tilde{\varepsilon} \right\}$  gilt  $|H_{i,j}^{(n)}| \leq |H_{i,l}^{(n)}|$ . Außerdem gilt generell  $\mathbb{1}_{A \cap B} \leq \mathbb{1}_A$ .
- d. Beide inneren Summanden konvergieren gemäß (G2').

Damit ist (G2) bewiesen und es folgt gemäß Lemma 2.1.1, dass

$$\sum_{j=1}^m c_j \sum_{i=1}^n \int H_{i,j}^{(n)}(s) dM_i(s) \implies N \left( 0, \sum_{j=1}^m \sum_{k=1}^m c_j c_k a_{jk} \right).$$

Dies ist aber genau die Verteilung von  $\sum_{j=1}^m c_j X_j$  für  $(X_1, \dots, X_m) \sim N(\mathbf{0}, A)$ . Nach dem Satz von Cramér-Wold ist das Lemma demnach bewiesen.  $\square$

## 2.3 Gleichmäßige Konvergenz des Varianzschätzers

Wir wollen zeigen, dass der Varianzschätzer  $\hat{\sigma}_x^2$  (siehe Satz 2.1.4, (c)) sogar gleichmäßig in Wahrscheinlichkeit konvergiert. Dazu sei folgende zusätzliche Annahme formuliert:

(K'') Es gilt (K'). Außerdem sei  $k^2$  Lipschitz-stetig, d.h. es existiert ein  $\tilde{C} > 0$ , sodass  $|k^2(u) - k^2(v)| \leq \tilde{C}|u - v|$  für alle  $u, v$ .

Wie zuvor sei  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_{d+1}$  ein  $(d+1)$ -dimensionales Teilintervall von  $[0, 1]^{d+1}$ , sodass jedes  $\mathcal{X}_i$  ein kompaktes Intervall ist.

### Satz 2.3.1

Unter den Annahmen (D'), (S'), (K'') und (B') gilt

$$\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^2 - \sigma_x^2| \xrightarrow{P} 0.$$

**Beweis:** Zunächst sei an die Definition von  $\hat{\sigma}_x^2$  aus Satz 2.1.4 (c) erinnert. Außerdem übernehmen wir von NL die Definition

$$\begin{aligned} \tilde{\sigma}_x^2 &:= \mathcal{C}_x^{-2} \frac{1}{n} b^{d+1} \sum_{i=1}^n \int K_b^2(x - X_i(s)) d\langle M_i \rangle(s) \\ &= \mathcal{C}_x^{-2} \frac{1}{n} b^{d+1} \sum_{i=1}^n \int K_b^2(x - X_i(s)) \alpha(X_i(s)) Y_i(s) ds \end{aligned}$$

und nutzen dann die Zerlegung

$$\begin{aligned} \sup_x |\hat{\sigma}_x^2 - \sigma_x^2| &\leq \sup_x |\hat{\sigma}_x^2 - \tilde{\sigma}_x^2| + \sup_x |\tilde{\sigma}_x^2 - \sigma_x^2| \\ &\leq \sup_x |\hat{\sigma}_x^2 - \tilde{\sigma}_x^2| + \sup_x |\tilde{\sigma}_x^2 - \mathbb{E}\tilde{\sigma}_x^2| + \sup_x |\mathbb{E}\tilde{\sigma}_x^2 - \sigma_x^2|. \end{aligned}$$

Es genügt die Konvergenz aller Terme auf der rechten Seite der Ungleichung zu zeigen. Da in NL im Beweis von Theorem 2 bereits  $\sup_x |\mathcal{C}_x - \varphi(x)| \xrightarrow{P} 0$  behandelt wird, ist es hinreichend, die Konvergenz von  $\sup_x |\mathcal{C}_x^2 \hat{\sigma}_x^2 - \mathcal{C}_x^2 \tilde{\sigma}_x^2|$ ,  $\sup_x |\mathcal{C}_x^2 \tilde{\sigma}_x^2 - \mathbb{E}\mathcal{C}_x^2 \tilde{\sigma}_x^2|$  und  $\sup_x |\mathbb{E}\mathcal{C}_x^2 \tilde{\sigma}_x^2 - \varphi^2(x) \sigma_x^2|$  zu zeigen. Dies liegt daran, dass aus  $f_n \rightarrow f$  gleichmäßig,  $g_n \rightarrow g$  gleichmäßig sowie  $\inf g > 0$  folgt, dass ebenfalls  $f_n/g_n \rightarrow f/g$  gleichmäßig.

Für die ersten beiden dieser Terme soll jeweils Lemma 2.1.3 genutzt werden. Da die punktweise Konvergenz schon in NL gezeigt wurde, beschränken wir uns auf den Nachweis von Bedingung (L2). In (L2) betrachten wir o.B.d.A. nur den Fall  $j = 1$ . Es

sei dafür  $x^* = (z^*, z_2, \dots, z_d, t)$ . Dafür gilt

$$\begin{aligned}
 & \mathbb{E} [\mathcal{C}_x^2 \hat{\sigma}_x^2 - \mathcal{C}_x^2 \tilde{\sigma}_x^2 - (\mathcal{C}_{x^*}^2 \hat{\sigma}_{x^*}^2 - \mathcal{C}_{x^*}^2 \tilde{\sigma}_{x^*}^2)]^2 \\
 & \stackrel{a.}{=} \mathbb{E} \left[ \frac{1}{n} b^{d+1} \sum_{i=1}^n \int \{K_b^2(x - X_i(s)) - K_b^2(x^* - X_i(s))\} dM_i(s) \right]^2 \\
 & \stackrel{b.}{=} \frac{1}{n^2} b^{2d+2} \sum_{i=1}^n \mathbb{E} \int \{K_b^2(x - X_i(s)) - K_b^2(x^* - X_i(s))\}^2 d\langle M_i \rangle(s) \\
 & \stackrel{c.}{=} \frac{1}{n} b^{2d+2} \int_{[0,1]^{d+1}} \{K_b^2(x - w) - K_b^2(x^* - w)\}^2 \alpha(w) \varphi(w) dw \\
 & \stackrel{d.}{=} \frac{1}{n} b^{-(d+1)} \int_{-x/b}^{(1-x)/b} \left\{ K^2(q) - K^2\left(q + \frac{x^* - x}{b}\right) \right\}^2 \alpha(x - bq) \varphi(x - bq) dq \\
 & \stackrel{e.}{\leq} \text{const} \cdot \frac{1}{n} b^{-(d+3)} |z^* - z_1|^2.
 \end{aligned} \tag{2.4}$$

- a. Definitionen von  $\hat{\sigma}_x^2$  und  $\tilde{\sigma}_x^2$  sowie die Definition von  $M_i$ .
- b. Wir nutzen, dass für allgemeine zentrierte Martingale  $X$  gilt, dass  $\mathbb{E}X(T)^2 = \mathbb{E}\langle X \rangle(T)$ , dass  $\int HdM$  ein derartiges Martingal ist sowie, dass  $\langle \int HdM \rangle = \int H^2 d\langle M \rangle$ , für Martingale  $M$ , die die Differenz zwischen einem Zählprozess und seinem Kompensator darstellen (siehe etwa Fleming u. Harrington (2011), Kapitel 2). Außerdem geht die Unabhängigkeit der Beobachtungen ein, die sicherstellt, dass wir die gemischten Terme ignorieren können, da die Summanden Erwartung null besitzen.
- c.  $d\langle M_i \rangle(s) = \alpha(X_i(s))Y_i(s)ds$  sowie Lemma 2.2.1.
- d. Substitution  $q = (x - w)/b$ .
- e. Wir stellen zunächst fest, dass aufgrund von (K'') und da sich  $x$  und  $x^*$  nur in der ersten Koordinate unterscheiden

$$\begin{aligned}
 \left| K^2(q) - K^2\left(q + \frac{x^* - x}{b}\right) \right| &= \left| k^2(q_1) - k^2\left(q_1 + \frac{z^* - z_1}{b}\right) \right| \cdot \prod_{j=2}^{d+1} k^2(q_j) \\
 &\leq \tilde{C} b^{-1} |z^* - z_1| \cdot \prod_{j=2}^{d+1} k^2(q_j).
 \end{aligned}$$

Setzt man dies ein, erhält man die Umformung, indem man erkennt, dass das verbleibende Integral beschränkt bleibt, da alle Integranden stetig sind und der Integrationsbereich kompakt bleibt (aufgrund des kompakten Trägers von  $K$ ).

Damit ist (L2) wegen (B') für  $\mathcal{C}_x^2 \hat{\sigma}_x^2 - \mathcal{C}_x^2 \tilde{\sigma}_x^2$  gezeigt.

Weiterhin erhalten wir

$$\begin{aligned}
 & \mathbb{E} \left[ \mathcal{C}_x^2 \tilde{\sigma}_x^2 - \mathbb{E} \mathcal{C}_x^2 \tilde{\sigma}_x^2 - (\mathcal{C}_{x^*}^2 \tilde{\sigma}_{x^*}^2 - \mathbb{E} \mathcal{C}_{x^*}^2 \tilde{\sigma}_{x^*}^2) \right]^2 \\
 &= \text{Var}(\mathcal{C}_x^2 \tilde{\sigma}_x^2 - \mathcal{C}_{x^*}^2 \tilde{\sigma}_{x^*}^2) \\
 &\stackrel{a.}{=} \text{Var} \left( \frac{1}{n} b^{d+1} \sum_{i=1}^n \int \{K_b^2(x - X_i(s)) - K_b^2(x^* - X_i(s))\} \alpha(X_i(s)) Y_i(s) ds \right) \\
 &\stackrel{b.}{\leq} \frac{1}{n} b^{2d+2} \int_{[0,1]^{d+1}} \{K_b^2(x - w) - K_b^2(x^* - w)\}^2 \alpha^2(w) \varphi(w) dw
 \end{aligned}$$

a. Definition von  $\tilde{\sigma}_x^2$  sowie  $d\langle M_i \rangle(s) = \alpha(X_i(s)) Y_i(s) ds$ .

b. Lemma 2.1.2.

Da der letzte Term bis auf den beschränkten Faktor  $\alpha(w)$  innerhalb des Integrals identisch (2.4) ist, ist (L2) auch für  $\mathcal{C}_x^2 \tilde{\sigma}_x^2 - \mathbb{E} \mathcal{C}_x^2 \tilde{\sigma}_x^2$  bewiesen.

Schließlich ist noch der Term  $|\mathbb{E} \mathcal{C}_x^2 \tilde{\sigma}_x^2 - \varphi^2(x) \sigma_x^2|$  zu untersuchen. Für diesen erhalten wir

$$\begin{aligned}
 |\mathbb{E} \mathcal{C}_x^2 \tilde{\sigma}_x^2 - \varphi^2(x) \sigma_x^2| &\stackrel{a.}{=} \left| b^{d+1} \int_{[0,1]^{d+1}} K_b^2(x - w) \alpha(w) \varphi(w) dw - \kappa_2^{d+1} \alpha(x) \varphi(x) \right| \\
 &\stackrel{b.}{=} \left| \int_{[-1,1]^{d+1}} K^2(q) \{ \alpha(x - bq) \varphi(x - bq) - \alpha(x) \varphi(x) \} dq \right|.
 \end{aligned}$$

a. Einsetzen der Definitionen sowie Lemma 2.2.1.

b. Dies gilt für hinreichend große  $n$ , also kleine  $b$ , wegen der Substitution  $q = (x - w)/b$  sowie der Tatsache, dass der Träger von  $K$  identisch  $[-1, 1]^{d+1}$  ist. Außerdem geht die Definition von  $\kappa_2$  ein.

Da der führende Faktor  $K^2$  des Integranden beschränkt ist, ergibt sich die auf  $\mathcal{X}$  gleichmäßige Konvergenz dieses Termes gegen null nun aus Lemma 2.3.2, welches unter diesem Beweis erläutert wird.

Damit ist der Beweis von Satz 2.3.1 abgeschlossen.  $\square$

### Lemma 2.3.2

Seien  $A \subset \mathbb{R}^d$  und  $Q \subset \mathbb{R}^d$  kompakt und  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  stetig. Sei  $b_n$  eine Nullfolge. Dann gilt

$$\sup_{x \in A} \left| \int_Q \{g(x) - g(x + b_n q)\} dq \right| \rightarrow 0.$$

**Beweis:** Mit  $C = \{x + bq | x \in A, q \in Q, b \in [0, 1]\}$  ist  $g$  gleichmäßig stetig auf  $C$ , da diese Menge kompakt ist. Das bedeutet, für jedes  $\varepsilon > 0$  existiert ein  $\delta > 0$ , sodass für alle  $x, y \in C$  gilt

$$|x - y| < \delta \Rightarrow |g(x) - g(y)| < \varepsilon,$$

wobei  $|\cdot|$  hier auch die euklidische Norm im  $\mathbb{R}^d$  bezeichnet. Seien  $\varepsilon > 0$  und  $\delta > 0$  gemäß dieser Definition. Da  $Q$  kompakt ist, existiert ein  $N$  derart, dass für alle  $n \geq N$

$$b_n \leq 1 \quad \text{und} \quad |b_n q| < \delta \quad \forall q \in Q.$$

Für alle  $n \geq N$  folgt nun

$$\sup_{x \in A} \left| \int_Q g(x) - g(x + b_n q) dq \right| \leq \sup_{x \in A} \int_Q \varepsilon dq = \varepsilon \cdot \lambda_d(Q),$$

wobei  $\lambda_d$  das  $d$ -dimensionale Lebesgue-Maß bezeichnet. Da  $\lambda_d(Q) < \infty$ , ist der Beweis abgeschlossen.  $\square$

## 2.4 Zwei Modellspezifikationstests

In diesem Kapitel sollen zwei Versionen eines Tests auf Modellspezifikation vorgestellt und behandelt werden, die auf dem Intensitätsschätzer von NL beruhen. Die Methodik der Tests stammt dabei von Gozalo (1993), der analoge Tests für den Fall nichtparametrischer Regression mit dem Nadaraya-Watson-Schätzer eingeführt hat. Unterschiedliche Fragen der Modellspezifikation können mit diesen Tests untersucht werden. Im späteren Verlauf werden wir insbesondere an einem Test auf Auslassen von Kovariablen interessiert sein. Aber auch das Testen von restriktiveren parametrischen, semiparametrischen oder strukturierten nichtparametrischen Modellformen ist möglich. Etwa könnte man untersuchen, ob das weit verbreitete semiparametrische Cox-Modell (Cox, 1972) oder ein multiplikatives bzw. additives nichtparametrisches Modell (siehe etwa Linton u. a., 2003) für die Daten geeignet ist.

Wir gehen weiterhin davon aus, dass wir uns im Modellrahmen von Kapitel 2.1.1 bewegen. Es sei  $\mathcal{B}$  eine Teilmenge der messbaren Funktionen, die von  $[0, 1]^{d+1}$  nach  $[0, \infty)$  abbilden. Die zu testende Nullhypothese lautet

**H<sub>0</sub>**: Es existiert eine Funktion  $\beta \in \mathcal{B}$ , sodass

$$P(\alpha(X_i(t)) = \beta(X_i(t)), t \in [0, 1]) = 1.$$

Dementsprechend lautet die Alternativhypothese

**H<sub>1</sub>**: Für alle Funktionen  $\beta \in \mathcal{B}$  gilt

$$P(\alpha(X_i(t)) = \beta(X_i(t)), t \in [0, 1]) < 1.$$



**Bemerkung:** Offensichtlich steuert die Wahl der Menge  $\mathcal{B}$  die Art des Tests. Für einen Test auf Auslassen der ersten Kovariaten wähle man etwa

$$\mathcal{B} = \left\{ \beta : [0, 1]^{d+1} \rightarrow [0, \infty) \mid \beta(x_1, x_2, \dots, x_{d+1}) \equiv \beta(x_2, \dots, x_{d+1}), x \in [0, 1]^{d+1} \right\}.$$

Für einen Test auf das Cox-Modell nutze man

$$\mathcal{B} = \left\{ \beta \mid \beta(z, t) = \lambda(t) \exp(\theta^T z), (z, t) \in [0, 1]^{d+1} \text{ für ein } \lambda : [0, 1] \rightarrow \mathbb{R}^+, \theta \in \mathbb{R}^d \right\}.$$

Wir betrachten ein fixes, endliches Gitter bestehend aus paarweise verschiedenen Punkten  $x^1, \dots, x^m \in (0, 1)^{d+1}$ . Weiterhin seien  $\hat{\alpha}$  der in Kapitel 2.1.2 definierte Schätzer sowie  $\hat{\sigma}_x^2$  wie in Satz 2.1.4. Darüber hinaus sei  $\hat{\beta}$  ein Schätzer für die Funktion  $\alpha$  unter der Nullhypothese. Wir definieren nun

$$S_n(x) := (nb^{d+1})^{\frac{1}{2}} \frac{\hat{\alpha}(x) - \hat{\beta}(x)}{\hat{\sigma}_x}$$

und damit unsere Teststatistik

$$T_n := \sum_{j=1}^m S_n(x^j)^2.$$

**Test (Gozalo-Typ I):** Lehne die Nullhypothese ab, falls die Teststatistik  $T_n$  einen größeren Wert aufweist als das  $(1 - \gamma)$ -Quantil der  $\chi^2$ -Verteilung mit  $m$  Freiheitsgraden.

■

Um zu zeigen, dass dessen asymptotisches Niveau  $\gamma$  ist, sind noch Annahmen an die Eigenschaften des Schätzers unter Nullhypothese notwendig. Dieser soll nämlich konsistent sein mit einer schnelleren Konvergenzgeschwindigkeit als der Schätzer für den allgemeinen Fall  $\hat{\alpha}$ .

(N) Für die reelle Zahlenfolge  $\{a_n^\beta\}$ , welche  $(nb^{d+1})^{1/2}/a_n^\beta \rightarrow 0$  erfülle, gelte unter  $\mathbf{H}_0$  für  $x \in \{x^1, \dots, x^m\}$

$$a_n^\beta \left( \hat{\beta}(x) - \alpha(x) \right) \xrightarrow{d} \xi,$$

wobei  $\xi$  eine Zufallsvariable mit Erwartungswert 0 und Varianz  $\sigma_\beta^2(x)$  ist, welche konsistent durch  $\hat{\sigma}_\beta^2(x)$  geschätzt werden kann.

Für eine Analyse des Tests unter  $\mathbf{H}_1$  können wir eine noch stärkere Annahme formulieren.

(A) Für die reelle Zahlenfolge  $\{a_n^\beta\}$ , welche  $(nb^{d+1})^{1/2}/a_n^\beta \rightarrow 0$  erfülle, und eine Funktion

$\beta_0 \in \mathcal{B}$  gelte für  $x \in \{x^1, \dots, x^m\}$

$$a_n^\beta \left( \hat{\beta}(x) - \beta_0(x) \right) \xrightarrow{d} \xi,$$

wobei  $\xi$  eine Zufallsvariable mit Erwartungswert 0 und Varianz  $\sigma_\beta^2(x)$  ist, welche konsistent durch  $\hat{\sigma}_\beta^2(x)$  geschätzt werden kann. Unter  $\mathbf{H}_0$  gilt  $\alpha = \beta_0$ .

**Satz 2.4.1**

Für jeden der Punkte  $x^1, \dots, x^m$  seien die Annahmen (D) und (S) erfüllt. Außerdem gelten (K), (B) sowie  $(\tilde{B})$ .

(a) Dann folgt unter (N) und  $\mathbf{H}_0$  für  $n \rightarrow \infty$

$$T_n \implies \chi_m^2.$$

(b) Es sei außerdem  $\{k_n\} = O((nb^{d+1})^\delta)$  eine reelle Zahlenfolge für ein  $\delta \in (0, 1)$ . Unter (A) und  $\mathbf{H}_1$  gilt, falls ein  $j$  existiert, sodass  $\alpha(x^j) \neq \beta_0(x^j)$  für  $n \rightarrow \infty$

$$P(T_n > k_n) \rightarrow 1.$$

(c) Es sei  $\{k_n\}$  wie in (b) und außerdem  $P(\hat{\beta} \in \mathcal{B}) = 1$ . Wenn nun wieder  $\mathbf{H}_1$  zusammen mit

$$\inf_{\beta \in \mathcal{B}} \max_{j=1, \dots, m} |\alpha(x^j) - \beta(x^j)| > 0,$$

gilt, so gilt ebenfalls für  $n \rightarrow \infty$

$$P(T_n > k_n) \rightarrow 1.$$

**Bemerkung:** Wir haben zwei Aussagen über das Verhalten des Tests unter  $\mathbf{H}_1$  formuliert. Satz 2.4.1 (b) folgt im Wesentlichen der in Gozalo (1993) gemachten Aussage. Häufig wird aber in der Literatur das Verhalten eines Schätzers für den Fall, dass die Modellannahmen nicht erfüllt sind, nicht analysiert. Ohne eine solche Analyse ist (b) aber nicht anwendbar. Daher beschreibt (c) eine alternative Bedingung, die die Konsistenz des Tests sicherstellt. Für den Fall eines Tests auf Auslassen der ersten Kovariaten wäre diese etwa erfüllt, falls unter den Gitterpunkten zwei Punkte  $x^1, x^2$  enthalten sind, die sich nur in der ersten Koordinate unterscheiden und an denen  $\alpha$  unterschiedliche Werte annimmt, d.h.

$$\begin{aligned} x_l^1 &= x_l^2, l = 2, \dots, d+1, \\ \alpha(x^1) &\neq \alpha(x^2). \end{aligned}$$

**Beweis:** (a) Wir betrachten zunächst für  $j = 1, \dots, m$

$$\begin{aligned} S_n(x^j) &= (nb^{d+1})^{\frac{1}{2}} \frac{\hat{\alpha}(x^j) - \hat{\beta}(x^j)}{\hat{\sigma}_{x^j}} \\ &= (nb^{d+1})^{\frac{1}{2}} \frac{\hat{\alpha}(x^j) - \alpha(x^j)}{\hat{\sigma}_{x^j}} + \frac{(nb^{d+1})^{\frac{1}{2}} \hat{\sigma}_{\beta}(x^j)}{a_n^{\beta}} a_n^{\beta} \frac{\alpha(x^j) - \hat{\beta}(x^j)}{\hat{\sigma}_{\beta}(x^j)}. \end{aligned} \quad (2.5)$$

Der zweite Summand konvergiert gemäß Annahme (N) gegen null in Wahrscheinlichkeit, da der vordere Bruch gegen null konvergiert und der restliche Term  $O_P(1)$  ist (siehe z.B. van der Vaart, 1998, Kapitel 2.2, für einen Überblick über die stochastische Landau- $O$ -Notation). Der vordere Term konvergiert gegen eine Standardnormalverteilung. Genauer gesagt, konvergiert sogar der Vektor der vorderen Terme (mit den Komponenten für  $j = 1, \dots, m$ ) Satz 2.2.3 zufolge in Verteilung gegen einen standardnormalverteilten  $m$ -dimensionalen Zufallsvektor. Somit gilt nach Slutzkys Lemma (siehe etwa van der Vaart, 1998, Lemma 2.8) dieselbe Konvergenz auch für den Vektor  $[S_n(x^j)]_{j=1, \dots, m}$ . Nach dem Continuous Mapping Theorem (siehe etwa van der Vaart, 1998, Theorem 2.3) folgt damit

$$T_n = \sum_{j=1}^m S_n(x^j)^2 \implies \chi_m^2.$$

(b) Wir erweitern zunächst (2.5) zu

$$\begin{aligned} S_n(x^j) &= (nb^{d+1})^{\frac{1}{2}} \frac{\hat{\alpha}(x^j) - \alpha(x^j)}{\hat{\sigma}_{x^j}} + \frac{(nb^{d+1})^{\frac{1}{2}} \hat{\sigma}_{\beta}(x^j)}{a_n^{\beta}} a_n^{\beta} \frac{\beta_0(x^j) - \hat{\beta}(x^j)}{\hat{\sigma}_{\beta}(x^j)} \\ &\quad + (nb^{d+1})^{\frac{1}{2}} \frac{\alpha(x^j) - \beta_0(x^j)}{\hat{\sigma}_{x^j}}. \end{aligned}$$

Die ersten beiden Summanden verhalten sich aufgrund von Annahme (A) genauso wie die beiden Summanden in (a) und sind somit insbesondere  $O_P(1)$ . Der Zähler des dritten Summanden ist annahmegemäß für zumindest ein  $j$  – wir bezeichnen es mit  $J$  – ungleich null, sodass der Term für  $j = J$  insgesamt mit Rate  $(nb^{d+1})^{1/2}$  gegen  $\pm\infty$  divergiert (man beachte hier Annahme (B)). Somit folgt für dieses  $J$ , dass  $P(|S_n(x^J)| > k_n^{1/2}) \rightarrow 1$ , da  $k_n^{1/2} = O((nb^{d+1})^{\delta/2})$  und  $\delta < 1$ . Damit ergibt sich

$$\begin{aligned} P(T_n > k_n) &= P\left(\sum_{j=1}^m S_n(x^j)^2 > k_n\right) \geq P(S_n(x^J)^2 > k_n) \\ &= P(|S_n(x^J)| > k_n^{1/2}) \rightarrow 1. \end{aligned}$$

(c) Wir stellen fest, dass

$$\begin{aligned} \max_{j=1,\dots,m} |S_n(x^j)| &\stackrel{a.}{\geq} \frac{(nb^{d+1})^{1/2}}{\hat{\sigma}_{x^j}} \left( \max_j |\alpha(x^j) - \hat{\beta}(x^j)| - \max_j |\alpha(x^j) - \alpha(x^j)| \right) \\ &\stackrel{b.}{\geq} \frac{(nb^{d+1})^{1/2}}{\hat{\sigma}_{x^j}} \inf_{\beta \in \mathcal{B}} \max_j |\alpha(x^j) - \beta(x^j)| - \frac{(nb^{d+1})^{1/2}}{\hat{\sigma}_{x^j}} \max_j |\alpha(x^j) - \alpha(x^j)| \end{aligned}$$

a. Wir fügen, ähnlich wie im Beweis von (a), die nahrhafte Null  $\pm\alpha(x^j)$  ein und nutzen dann die Regel  $\max_x |f(x) + g(x)| \geq \max_x |f(x)| - \max_x |g(x)|$ .

b. Da  $\hat{\beta} \in \mathcal{B}$  können wir den ersten Summanden durch das Infimum über alle  $\beta \in \mathcal{B}$  abschätzen.

Der zweite (negative) Summand ist  $O_P(1)$ . Der erste divergiert mit Rate  $(nb^{d+1})^{1/2}$  gegen  $\infty$ , weswegen der gesamte Term mit dieser Rate divergiert. Somit folgt  $P(\max_j |S_n(x^j)| > k_n^{1/2}) \rightarrow 1$  und damit wiederum

$$\begin{aligned} P(T_n > k_n) &= P\left(\sum_{j=1}^m S_n(x^j)^2 > k_n\right) \geq P\left(\max_j S_n(x^j)^2 > k_n\right) \\ &= P\left(\max_j S_n(x^j) > k_n^{1/2}\right) \rightarrow 1. \end{aligned}$$

□

Teile (b) und (c) des Satzes 2.4.1 weisen die Konsistenz des beschriebenen Tests unter verschiedenen Szenarien nach. Jeweils werden aber Bedingungen an die deterministischen Punkte  $x^j$  gestellt, welche nicht überprüfbar sind, da die tatsächliche Funktion  $\alpha$  und gegebenenfalls  $\beta_0$  unbekannt sind. Ein Lösungsansatz kann sein, die  $x^j$  zufällig zu bestimmen und die Anzahl  $m$  mit wachsendem Stichprobenumfang  $n$  ebenfalls in geeigneter Geschwindigkeit zu erhöhen. So werden asymptotisch zumindest mit Wahrscheinlichkeit eins passende  $x^j$  gewählt. Gozalo (1993) führt diesen Gedanken noch einen Schritt weiter und ergänzt ihn um einen Suchmechanismus. Wir wollen ebenfalls eine derartige Version für unseren Fall konstruieren.

Es seien dazu  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_{d+1}$  ein  $(d+1)$ -dimensionales Teilintervall von  $[0, 1]^{d+1}$ , sodass jedes  $\mathcal{X}_i$  ein kompaktes Intervall ist, und  $f_{\tilde{X}}$  die Dichte einer stetigen Verteilung auf dieser Menge. Es sei  $\{\tilde{X}^j\}_{j \in \mathbb{N}}$  eine Folge unabhängiger, identisch gemäß  $f_{\tilde{X}}$  verteilter Zufallsvektoren, welche zudem unabhängig von den Daten sei. Darüber hinaus sei  $\{m_n\}_{n \in \mathbb{N}}$  eine Folge natürlicher Zahlen mit

$$m_n \rightarrow \infty, \quad m_n = o((nb^{d+1})^\delta) \quad (2.6)$$

für ein  $\delta \in (0, 1)$  und  $n \rightarrow \infty$ . Nun definieren wir für ein  $\eta > 0$

$$J := \arg \max_{1 \leq j \leq m_n} \{S_n(\tilde{X}^j)^2 - \eta(nb^{d+1})^\delta \mathbb{1}_{\{j > 1\}}\} \quad (2.7)$$

und

$$\hat{X} = \tilde{X}^J.$$

**Test (Gozalo-Typ II):** Lehne die Nullhypothese ab, falls die Teststatistik  $S_n(\hat{X})^2$  einen größeren Wert aufweist als das  $(1 - \gamma)$ -Quantil der  $\chi^2$ -Verteilung mit einem Freiheitsgrad. ■

Dessen asymptotische Eigenschaften fassen wir in Satz 2.4.2 zusammen. Wir formulieren hier eine weitere Annahme für die Untersuchung des Tests unter der Alternativhypothese.

(A') Es existiert eine Funktion  $\beta_0 \in \mathcal{B}$ , für welche für  $n \rightarrow \infty$  gilt

$$\sup_{x \in \mathcal{X}} |\hat{\beta}(x) - \beta_0(x)| \rightarrow_P 0.$$

### Satz 2.4.2

Es seien  $(D')$ ,  $(K'')$ ,  $(B')$  und  $(\tilde{B})$  erfüllt sowie  $(S)$  für alle  $x \in \mathcal{X}$ .

(a) Unter  $\mathbf{H}_0$  und  $(N)$  folgen für  $n \rightarrow \infty$

$$P(J = 1) \rightarrow 1$$

und

$$S_n(\hat{X})^2 \Longrightarrow \chi_1^2.$$

(b) Es gelten  $\mathbf{H}_1$ , Annahme  $(A')$  sowie

$$P(\alpha(\tilde{X}) \neq \beta_0(\tilde{X})) > 0. \quad (2.8)$$

Für  $\{k_n\}_n = O((nb^{d+1})^\delta)$  folgt dann für  $n \rightarrow \infty$

$$P(S_n(\hat{X})^2 > k_n) \rightarrow 1.$$

(c) Es sei  $\{k_n\}_n$  wie in (b) und außerdem  $P(\hat{\beta} \in \mathcal{B}) = 1$ . Wenn nun wieder  $\mathbf{H}_1$  gilt sowie ein  $\varepsilon > 0$  derart existiert, dass

$$\inf_{\beta \in \mathcal{B}} P(|\alpha(\tilde{X}) - \beta(\tilde{X})| > \varepsilon) =: p > 0,$$

so folgt ebenfalls für  $n \rightarrow \infty$

$$P(S_n(\hat{X})^2 > k_n) \rightarrow 1.$$

**Bemerkung:** Erneut formulieren wir zwei Varianten von Aussagen über das Verhalten des Tests unter der Alternativhypothese. Die Gründe sind die gleichen wie bei Satz 2.4.1. Wir betrachten wieder das Beispiel eines Tests auf Auslassen der ersten Kovariablen. Die Voraussetzung von (c) wäre etwa erfüllt, wenn  $\alpha$  stetig ist, wir gleichmäßige Stetigkeit  $\mathcal{B}$  annehmen sowie  $x, y \in \mathcal{X}$  existieren, die sich nur in der ersten Koordinate unterscheiden, in denen  $\alpha$  aber unterschiedliche Werte annimmt und die innerhalb einer konvexen Menge liegen, auf der  $f_{\tilde{X}}$  von null weg beschränkt ist, d.h.

$$\begin{aligned} x, y &\in \mathcal{K} \subset \mathcal{X}, \quad \mathcal{K} \text{ konvex,} \\ \exists \varepsilon > 0 : f_{\tilde{X}}|_{\mathcal{K}} &> \varepsilon, \\ x_j &= y_j, j = 2, \dots, d+1, \quad \text{und} \\ \alpha(x) &\neq \alpha(y). \end{aligned}$$

Für den Beweis von Satz 2.4.2 formulieren wir ein Lemma, dessen Beweis wir an das Ende dieses Kapitels schieben.

### Lemma 2.4.3

Unter  $\mathbf{H}_0$  sowie den Annahmen  $(D')$ ,  $(K'')$ ,  $(B')$ ,  $(\tilde{B})$  sowie  $(S)$  für alle  $x \in \mathcal{X}$  gilt

$$\frac{1}{m_n} \sum_{j=1}^{m_n} S_n(\tilde{X}^j)^2 = O_P(1).$$

**Beweis Satz 2.4.2:** (a) Für alle  $\mu > 0$  gilt

$$P\left(\max_{j=1, \dots, m_n} S_n(\tilde{X}^j)^2 \geq \mu (nb^{d+1})^\delta\right) \leq P\left(\frac{1}{m_n} \sum_{j=1}^{m_n} S_n(\tilde{X}^j)^2 \geq \mu \frac{(nb^{d+1})^\delta}{m_n}\right) \xrightarrow{n \rightarrow \infty} 0.$$

Zunächst wird hier das Maximum durch die Summe abgeschätzt und durch  $m_n$  geteilt. Die Konvergenz ergibt sich mit Lemma 2.4.3, da der Quotient  $(nb^{d+1})^\delta/m_n$  gegen unendlich divergiert. Denn allgemein gilt für  $X_n = O_P(1)$ ,  $a_n \rightarrow \infty$ , dass für jedes  $\varepsilon > 0$  Größen  $M > 0$ ,  $N_1 > 0$  und  $N_2 > 0$  derart existieren, dass für alle  $n \geq N_1$  gilt  $P(|X_n| > M) < \varepsilon$ , und für alle  $n \geq N_2$  gilt  $a_n > M$ . Daher gilt dann für alle  $n \geq N_1 \vee N_2$ , dass  $P(|X_n| > a_n) < \varepsilon$ .

Nun ergibt sich der erste Teil von (a) durch

$$\begin{aligned}
 P(J \neq 1) &= P(\exists j \in \{2, \dots, m_n\} : S_n(\tilde{X}^j)^2 - S_n(\tilde{X}^1)^2 > \eta(nb^{d+1})^\delta) \\
 &\leq P(\exists j \in \{1, \dots, m_n\} : S_n(\tilde{X}^j)^2 > \eta(nb^{d+1})^\delta) \\
 &\leq P\left(\max_{j=1, \dots, m_n} S_n(\tilde{X}^j)^2 \geq \eta(nb^{d+1})^\delta\right) \xrightarrow{n \rightarrow \infty} 0.
 \end{aligned}$$

Der zweite Teil ergibt sich, weil  $S_n(\tilde{X}^1)^2 \Rightarrow \chi_1^2$  (siehe dafür Theorem 2.3 von Gozalo (1993), welches auf einen Konvergenzsatz von Bierens (1984) zurückgeht).

(b) Wir bemerken zunächst, dass

$$\begin{aligned}
 &\max_j |S_n(\tilde{X}^j)| \\
 &\stackrel{a.}{=} \max_j \frac{(nb^{d+1})^{1/2}}{\hat{\sigma}_{\tilde{X}^j}} \left| \hat{\alpha}(\tilde{X}^j) - \alpha(\tilde{X}^j) + \alpha(\tilde{X}^j) - \beta_0(\tilde{X}^j) + \beta_0(\tilde{X}^j) - \hat{\beta}(\tilde{X}^j) \right| \\
 &\stackrel{b.}{\geq} \frac{(nb^{d+1})^{1/2}}{\max_j \hat{\sigma}_{\tilde{X}^j}} \left( \max_j |\alpha(\tilde{X}^j) - \beta_0(\tilde{X}^j)| - \max_j |\hat{\alpha}(\tilde{X}^j) - \alpha(\tilde{X}^j)| - \max_j |\beta_0(\tilde{X}^j) - \hat{\beta}(\tilde{X}^j)| \right)
 \end{aligned} \tag{2.9}$$

a. Wir fügen der Definition von  $S_n$  zwei nahrhafte Nullen hinzu.

b. Wir nutzen die für Normen  $\|\cdot\|$  gültige Ungleichung  $\|x + y\| \geq \|x\| - \|y\|$ .

Wir erhalten weiterhin

$$\begin{aligned}
 P(S_n(\hat{X})^2 > k_n) &\stackrel{a.}{\geq} P\left(\max_{j=1, \dots, m_n} |S_n(\tilde{X}^j)| > (\eta(nb^{d+1})^\delta + k_n)^{1/2}\right) \\
 &\stackrel{b.}{\geq} P\left(\max_j |\alpha(\tilde{X}^j) - \beta_0(\tilde{X}^j)| > \max_j \hat{\sigma}_{\tilde{X}^j} \left(\frac{\eta(nb^{d+1})^\delta + k_n}{nb^{d+1}}\right)^{1/2} \right. \\
 &\quad \left. + \max_j |\hat{\alpha}(\tilde{X}^j) - \alpha(\tilde{X}^j)| + \max_j |\beta_0(\tilde{X}^j) - \hat{\beta}(\tilde{X}^j)| \right) \\
 &\stackrel{c.}{\rightarrow} 1.
 \end{aligned}$$

a. Gemäß der Definition von  $\hat{X}$  gilt  $S_n(\hat{X})^2 \geq \max_j S_n(\tilde{X}^j)^2 - \eta(nb^{d+1})^\delta$ . Wir setzen dies ein, schieben einen Term auf die andere Seite der Ungleichung und ziehen die Wurzel.

b. Wir setzen Formel (2.9) ein.

c. Alle Terme der rechten Seite der Ungleichung innerhalb der Wahrscheinlichkeit konvergieren in Wahrscheinlichkeit gegen null: der erste, weil  $\hat{\sigma}$  gleichmäßig in  $x$  konvergiert und  $k_n = O((nb^{d+1})^\delta)$  (man beachte Annahme (B')), welche (B)

impliziert), der zweite nach Satz 2.1.5 (man beachte, dass Annahme (S) stärker als (S') ist) und der dritte nach Annahme (A'). Die linke Seite ist aber wegen (2.8) asymptotisch von null weg beschränkt.

(c) Aufgrund einer zu der im Beweis von (b) analogen Argumentation gilt

$$\begin{aligned}
 & P(S_n(\hat{X})^2 > k_n) \\
 & \geq P\left(\max_j |\alpha(\tilde{X}^j) - \hat{\beta}(\tilde{X}^j)| > \max_j \hat{\sigma}_{\tilde{X}^j} \left(\frac{\eta(nb^{d+1})^\delta + k_n}{nb^{d+1}}\right)^{1/2} + \max_j |\hat{\alpha}(\tilde{X}^j) - \alpha(\tilde{X}^j)|\right).
 \end{aligned} \tag{2.10}$$

Aus ebenso analogen Gründen folgt auch, dass die rechte Seite der Ungleichung innerhalb der Wahrscheinlichkeit stochastisch gegen null konvergiert. Weiter ergibt sich für das  $\varepsilon$  aus der Voraussetzung des Satzes

$$\begin{aligned}
 P\left(\max_j |\alpha(\tilde{X}^j) - \hat{\beta}(\tilde{X}^j)| > \varepsilon\right) &= \int P\left(\max_j |\alpha(\tilde{X}^j) - \hat{\beta}(\tilde{X}^j)| > \varepsilon \mid \hat{\beta} = \beta\right) dP_{\hat{\beta}}(\beta) \\
 &= \int P\left(\max_j |\alpha(\tilde{X}^j) - \beta(\tilde{X}^j)| > \varepsilon\right) dP_{\hat{\beta}}(\beta),
 \end{aligned}$$

wobei die zweite Identität die Unabhängigkeit von  $\tilde{X}^j$  und den Daten ausnutzt. Außerdem gilt

$$\begin{aligned}
 P\left(\max_j |\alpha(\tilde{X}^j) - \beta(\tilde{X}^j)| > \varepsilon\right) &= 1 - P(|\alpha(\tilde{X}) - \beta(\tilde{X})| > \varepsilon)^{m_n} \\
 &\geq 1 - \sup_{\beta \in \mathcal{B}} P(|\alpha(\tilde{X}) - \beta(\tilde{X})| > \varepsilon)^{m_n} \\
 &= 1 - (1 - p)^{m_n} \rightarrow 1,
 \end{aligned}$$

mit  $p > 0$  aus der Voraussetzung des Satzes. Insgesamt erhalten wir so, dass die linke Seite innerhalb der Wahrscheinlichkeit in (2.10) asymptotisch von null weg beschränkt bleibt und somit (2.10) gegen eins konvergiert.  $\square$

Was bleibt, ist, Lemma 2.4.3 zu beweisen.

**Beweis Lemma 2.4.3:** Es genügt zu zeigen, dass die Aussage gilt, wenn wir  $S_n$  durch den ersten der Summanden folgender Zerlegung ersetzen, da  $\hat{\beta}$  unter  $\mathbf{H}_0$  asymptotisch näher an  $\alpha$  liegt als  $\hat{\alpha}$ .

$$S_n(\tilde{X}) = (nb^{d+1})^{\frac{1}{2}} \left\{ \left( \frac{\hat{\alpha}(\tilde{X}) - \alpha(\tilde{X})}{\hat{\sigma}_{\tilde{X}}} \right) + \left( \frac{\alpha(\tilde{X}) - \hat{\beta}(\tilde{X})}{\hat{\sigma}_{\tilde{X}}} \right) \right\}$$



Diesen Summanden nennen wir  $S'_n$  und können ihn wie folgt darstellen (siehe Kapitel 2.1):

$$\begin{aligned} S'_n(\tilde{X}) &:= (nb^{d+1})^{\frac{1}{2}} \left( \frac{\hat{\alpha}(\tilde{X}) - \alpha(\tilde{X})}{\hat{\sigma}_{\tilde{X}}} \right) \\ &= \frac{1}{\hat{\sigma}_{\tilde{X}}} \frac{1}{\mathcal{C}_{\tilde{X}}} (nb^{d+1})^{\frac{1}{2}} \left\{ \frac{1}{n} \sum_{i=1}^n \int K_b(\tilde{X} - X_i(s)) dM_i(s) \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=1}^n \int K_b(\tilde{X} - X_i(s)) [\alpha(X_i(s)) - \alpha(\tilde{X})] Y_i(s) ds \right\} \\ &=: \frac{1}{\hat{\sigma}_{\tilde{X}}} \frac{1}{\mathcal{C}_{\tilde{X}}} \{R_{n1}(\tilde{X}) + R_{n2}(\tilde{X})\} =: \frac{1}{\hat{\sigma}_{\tilde{X}}} \frac{1}{\mathcal{C}_{\tilde{X}}} R_n(\tilde{X}) \end{aligned}$$

Wir werden zunächst nur die ersten beiden Faktoren  $1/\hat{\sigma}_{\tilde{X}}$  sowie  $1/\mathcal{C}_{\tilde{X}}$  betrachten und zeigen, dass es genügt, den übrigen Term  $R_n(\tilde{X})$  isoliert zu untersuchen. Dies fassen wir in folgender Behauptung zusammen:

*Behauptung 1:* Falls  $1/m_n \cdot \sum_{j=1}^{m_n} R_n(\tilde{X}^j)^2 = O_P(1)$ , so folgt auch  $1/m_n \cdot \sum_{j=1}^{m_n} S'_n(\tilde{X}^j)^2 = O_P(1)$ .

*Beweis Behauptung 1:* Gemäß Satz 2.3.1 wissen wir, dass  $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x - \sigma_x| \rightarrow_P 0$ , und bereits in NL wird erwähnt, dass  $\sup_{x \in \mathcal{X}} |\mathcal{C}_x - \varphi(x)| \rightarrow_P 0$ . Damit folgt ebenfalls, dass  $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^2 \mathcal{C}_x^2 - \sigma_x^2 \varphi(x)^2| \rightarrow_P 0$ . Es seien  $s := \inf_x \sigma_x^2 \varphi(x)^2 / 2 > 0$  und

$$A_n := \left\{ \inf_{j=1, \dots, n} \hat{\sigma}_{\tilde{X}^j}^2 \mathcal{C}_{\tilde{X}^j}^2 < s \right\} \subset \left\{ \inf_{x \in \mathcal{X}} \hat{\sigma}_x^2 \mathcal{C}_x^2 < s \right\} \subset \left\{ \sup_{x \in \mathcal{X}} |\hat{\sigma}_x^2 \mathcal{C}_x^2 - \sigma_x^2 \varphi(x)^2| > s \right\}.$$

Dann gilt  $P(A_n) \rightarrow 0, n \rightarrow \infty$ . Gemäß der Annahme aus Behauptung 1 existiert für jedes  $\varepsilon > 0$  ein  $M > 0$  derart, dass  $P(1/m_n \cdot \sum_{j=1}^{m_n} R_n(\tilde{X}^j)^2 > s \cdot M) < \varepsilon, n \in \mathbb{N}$ . Damit folgt

$$\begin{aligned} &P \left( \frac{1}{m_n} \sum_{j=1}^{m_n} S'_n(\tilde{X}^j)^2 > M \right) \\ &\leq P \left( \frac{1}{m_n} \sum_{j=1}^{m_n} R_n(\tilde{X}^j)^2 > s \cdot M, A_n^c \right) + P \left( \frac{1}{m_n} \sum_{j=1}^{m_n} S'_n(\tilde{X}^j)^2 > M, A_n \right). \end{aligned}$$

Da der erste Summand kleiner als  $\varepsilon$  ist und der zweite gegen null konvergiert, ist die Behauptung gezeigt.

*Behauptung 2:* Falls  $\sup_x \mathbb{E} R_{nl}(x)^4 = O(1), l = 1, 2$ , so gilt auch  $1/m_n \cdot \sum_{j=1}^{m_n} R_n(\tilde{X}^j)^2 = O_P(1)$ .

*Beweis Behauptung 2:* Zunächst bemerken wir, dass

$$\begin{aligned}\mathbb{E}R_{nl}(\tilde{X})^4 &= \int \mathbb{E}[R_{nl}(\tilde{X})^4 | \tilde{X} = x] dP_{\tilde{X}}(x) = \int \mathbb{E}[R_{nl}(x)^4] dP_{\tilde{X}}(x) \\ &\leq \int \sup_x \mathbb{E}R_{nl}(x)^4 dP_{\tilde{X}}(x) = \sup_x \mathbb{E}R_{nl}(x)^4 = O(1), \quad l = 1, 2,\end{aligned}$$

wobei für die zweite Identität die stochastische Unabhängigkeit von  $\tilde{X}$  von den Daten genutzt wird. Mit Hilfe der Minkowski-Ungleichung lässt sich leicht einsehen, dass damit auch  $\mathbb{E}R_n(\tilde{X})^4 = O(1)$  einhergeht. Dies wiederum zieht sofort nach sich, dass  $\mathbb{E}R_n(\tilde{X})^2 = O(1)$  und  $\text{Var}R_n(\tilde{X})^2 = O(1)$ . Da die Summanden identisch (wenn auch nicht notwendigerweise unabhängig) verteilt sind, folgt damit auch die Beschränktheit der Erwartungswerte und Varianzen von  $1/m_n \cdot \sum_{j=1}^{m_n} R_n(\tilde{X}^j)^2$  und damit mit der Ungleichung von Chebychev unsere Behauptung 2.

Jede der Zufallsvariablen  $R_{nl}(x)$  hat die Form  $1/n \sum_{i=1}^n H_{ni}$  für gewisse Zufallsvariablen  $H_{ni}$ . Es lohnt sich daher, einen Blick auf diese Struktur zu werfen, um das Resultat für beide Terme gleichermaßen nutzen zu können.

*Behauptung 3:* Für  $n \in \mathbb{N}$  seien  $H_{ni}, i = 1, \dots, n$ , unabhängig identisch verteilte Zufallsvariablen. Falls  $\mathbb{E}H_{ni} = O((nb^{d+1})^{-1/2})$ ,  $\text{Var}H_{n1} = O(b^{-d-1})$  und  $\mathbb{E}H_{n1}^4 = O(b^{-3d-3})$ , so folgt

$$\mathbb{E} \left[ (nb^{d+1})^{1/2} 1/n \sum_{i=1}^n H_{ni} \right]^4 = O(1).$$

Falls die Summanden stattdessen stochastische Prozesse abhängig von einem Parameter  $x$  sind und die Voraussetzungen gleichmäßig in  $x$  erfüllt sind, so gilt auch die Folgerung gleichmäßig in  $x$ . Gleichmäßigkeit bedeutet hier, dass Suprema von Erwartungswerten betrachtet werden, nicht Erwartungswerte von Suprema.

*Beweis Behauptung 3:* Zunächst lässt sich mit der Hölder-Ungleichung leicht erkennen, dass für Folgen von Zufallsvariablen  $X_n, n \in \mathbb{N}$ , gilt

$$\begin{aligned}\mathbb{E}X_n^4 = O(1) &\Leftrightarrow \forall a_n = O(1) : \mathbb{E}[X_n - a_n]^4 = O(1) \\ &\Leftrightarrow \exists a_n = O(1) : \mathbb{E}[X_n - a_n]^4 = O(1).\end{aligned}\tag{2.11}$$

Insbesondere kann  $a_n = \mathbb{E}X_n$  eingesetzt werden, falls diese Folge beschränkt ist. Es genügt also zu zeigen, dass  $\mathbb{E}[(nb^{d+1})^{1/2} 1/n \sum_{i=1}^n (H_{ni} - \mathbb{E}H_{ni})]^4 = O(1)$  und dass  $(nb^{d+1})^{1/2} \mathbb{E}H_{ni} = O(1)$ . Letzteres ist gemäß der Annahmen der Fall. Für ersteren

Term erhalten wir dank der unabhängig identischen Verteilung von  $H_{ni}, i = 1, \dots, n$ ,

$$\begin{aligned} & \mathbb{E} \left[ (nb^{d+1})^{1/2} 1/n \sum_{i=1}^n (H_{ni} - \mathbb{E}H_{ni}) \right]^4 \\ &= (nb^{d+1})^2 \left\{ \frac{1}{n^3} \mathbb{E} [H_{ni} - \mathbb{E}H_{ni}]^4 + \frac{3(n-1)}{n^3} \text{Var}[H_{n1}]^2 \right\} \\ &= O\left((nb^{d+1})^{-1}\right) + O(1) = O(1) \end{aligned}$$

Die Abschätzungen der beiden Summanden folgen aus den Annahmen, wobei für den ersten Summanden auch relevant ist, dass  $\mathbb{E}H_{ni}$  beschränkt bleibt. Die in  $x$  gleichmäßige Version folgt auf identischem Wege, indem vor die entsprechenden Erwartungswerte  $\sup_x$  geschrieben wird. Damit ist Behauptung 3 gezeigt.

Wir werden nun die Zufallsvariablen  $R_{nl}, l = 1, 2$ , einzeln untersuchen und damit den Beweis des Lemmas abschließen. Wir beginnen mit dem einfacher zu handhabenden Term  $R_{n2}$ .

*Behauptung 4:*  $\sup_x \mathbb{E}R_{n2}(x)^4 = O(1)$ .

*Beweis Behauptung 4:* In der Notation von Behauptung 3 ist hier  $H_{ni}(x) = \int K_b(x - X_i(s)) [\alpha(X_i(s)) - \alpha(x)] Y_i(s) ds$ . Aus den Herleitungen von NL im Beweis von deren Theorem 1(b) ist leicht zu erkennen, dass

$$\begin{aligned} \sup_x \mathbb{E}H_{ni}(x) &= O(b^2) \text{ und} \\ \sup_x \text{Var}H_{ni}(x) &= O(b^{-d-1}), \end{aligned}$$

wobei aus ersterer Aussage mit Annahme ( $\tilde{B}$ ) auch folgt, dass  $\sup_x \mathbb{E}H_{ni}(x) = O((nb^{d+1})^{-1/2})$ . Schließlich erhalten wir

$$\begin{aligned} \sup_x \mathbb{E}H_{ni}(x)^4 &\stackrel{a.}{\leq} \sup_x \int K_b^4(x-w) [\alpha(w) - \alpha(x)]^4 \varphi(w) dw \\ &\stackrel{b.}{=} \sup_x b^{-3d-3} \int_{[-1,1]^{d+1}} K^4(q) [\alpha(x-bq) - \alpha(x)] \varphi(x-bq) dq \\ &\stackrel{c.}{=} O(b^{-3d-3}). \end{aligned} \tag{2.12}$$

a. Zunächst liefert die Jensensche Ungleichung

$$\begin{aligned} & \left( \int K_b(x - X_i(s)) [\alpha(X_i(s)) - \alpha(x)] Y_i(s) ds \right)^4 \\ & \leq \int K_b^4(x - X_i(s)) [\alpha(X_i(s)) - \alpha(x)]^4 Y_i(s) ds. \end{aligned}$$

Im Anschluss wenden wir Lemma 2.2.1 an.

b. Substitution  $q = (x - w)/b$  sowie die Feststellung, dass der Integrationsbereich

auf das Einheitsintervall beschränkt bleibt, da  $K$  nur hier positive Werte besitzt.

- c. Wir können den Term abschätzen, indem wir das Supremum in das Integral hineinziehen. Die Integranden sind stetige Funktionen, der Integrationsbereich kompakt, sodass das Integral beschränkt bleibt.

Durch Anwendung von Behauptung 3 ist Behauptung 4 gezeigt.

*Behauptung 5:*  $\sup_x \mathbb{E} R_{n1}(x)^4 = O(1)$ .

*Beweis Behauptung 5:*  $R_{n1}(x) = \frac{1}{n} \sum_{i=1}^n \int K_b(x - X_i(s)) dM_i(s)$  ist ein Martingal für jedes  $x$ . Somit ist der Erwartungswert für alle  $x$  identisch null. Für die Varianzen gilt

$$\begin{aligned} \sup_x \text{Var} \int K_b(x - X_i(s)) dM_i(s) &\stackrel{a.}{=} \sup_x \mathbb{E} \int K_b^2(x - X_i(s)) \alpha(X_i(s)) Y_i(s) ds \\ &\stackrel{b.}{=} b^{-d-1} \sup_x \int_{[-1,1]^{d+1}} K^2(q) \alpha(x - bq) \varphi(x - bq) dq \\ &\stackrel{c.}{=} O(b^{-d-1}). \end{aligned}$$

- a. Wir nutzen, dass  $d\langle M_i \rangle(s) = \alpha(X_i(s)) Y_i(s) ds$  ist.  
 b. Lemma 2.2.1 und Substitution  $q = (x - w)/b$ .  
 c. Wieder sind die Integranden und der Integrationsbereich beschränkt.

Für das Abschätzen der vierten Momente nutzen wir folgende Aufspaltung:

$$\begin{aligned} \int K_b(x - X_i(s)) dM_i(s) &= \int K_b(x - X_i(s)) dN_i(s) - \int K_b(x - X_i(s)) \alpha(X_i(s)) Y_i(s) ds \\ &\leq \int K_b(x - X_i(s)) dN_i(s) + \int K_b(x - X_i(s)) \alpha(X_i(s)) Y_i(s) ds \end{aligned}$$

Es genügt, dass die vierten Momente beider Summanden einzeln der Anforderung von Behauptung 3 gehorchen. Für den zweiten Summanden folgt die Argumentation derjenigen aus (2.12). Es sei nun  $\tilde{\alpha} = \sup_x \alpha(x) < \infty$  gemäß (S) unter Ausnutzung der Kompaktheit von  $\mathcal{X}$ . Es sei weiterhin  $\tilde{N}$  ein homogener Poisson-Prozess mit Intensität  $\tilde{\alpha}$ , welcher stochastisch unabhängig von allen Daten ist. Da der Intensitätsprozess von  $N_i$ , also  $\alpha(X_i(s)) Y_i(s)$ , zu jedem Zeitpunkt kleiner ist als der Intensitätsprozess von  $\tilde{N}$ , also  $\tilde{\alpha}$ , folgt offenbar

$$\begin{aligned} \mathbb{E} \left[ \int K_b(x - X_i(s)) dN_i(s) \right]^4 &= \mathbb{E} \left[ \int K_b(x - X_i(s)) Y_i(s) dN_i(s) \right]^4 \\ &\leq \mathbb{E} \left[ \int K_b(x - X_i(s)) Y_i(s) d\tilde{N}(s) \right]^4, \end{aligned}$$

wobei die erste Gleichung gilt, da  $N_i$  nur an Punkten  $s$  springen kann, in denen  $Y_i(s) = 1$  ist. Nun folgt mit der Hölder-Ungleichung

$$\left[ \int K_b(x - X_i(s)) Y_i(s) d\tilde{N}(s) \right]^4 \leq \tilde{N}(1)^3 \cdot \int K_b^4(x - X_i(s)) Y_i(s) d\tilde{N}(s).$$

Wir wollen diesen Term nun zunächst bedingt dem Ereignis  $\{\tilde{N}(1) = k\}$  betrachten für  $k \in \mathbb{N}$ . Es seien dazu  $T_1, \dots, T_k$  die Sprungzeitpunkte von  $\tilde{N}$ . Diese sind, bedingt auf  $\{\tilde{N}(1) = k\}$ , unabhängig identisch gleichverteilt auf dem Einheitsintervall. Wir erhalten

$$\begin{aligned} & \mathbb{E} \left[ \tilde{N}(1)^3 \cdot \int K_b^4(x - X_i(s)) Y_i(s) d\tilde{N}(s) \middle| \tilde{N}(1) = k \right] \\ & \stackrel{a.}{=} k^3 \sum_{l=1}^k \int \mathbb{E} [K_b^4(x - X_i(t_l)) Y_i(t_l) | T_l = t_l] dP_{T_l}(t_l) \\ & \stackrel{b.}{\leq} k^3 b^{-4d-4} C_1 \sum_{l=1}^k \int \mathbb{1}_{\{|t-t_l|<b\}} \mathbb{E} [\mathbb{1}_{\{|z_j - Z_{ij}(t_l)|<b, j=1, \dots, d\}} Y_i(t_l)] dP_{T_l}(t_l) \\ & \stackrel{c.}{\leq} k^3 b^{-4d-4} b^d C_2 \sum_{l=1}^k \int \mathbb{1}_{\{|t-t_l|<b\}} dP_{T_l}(t_l) \\ & \stackrel{d.}{\leq} k^{3+1} b^{-3d-3} C_3 = k^4 O(b^{-3d-3}). \end{aligned}$$

- a. Auf dem Ereignis  $\{\tilde{N}(1) = k\}$  lässt sich  $\int K_b^4(x - X_i(s)) Y_i(s) d\tilde{N}(s)$  als  $\sum_{l=1}^k K_b^4(x - X_i(T_l)) Y_i(T_l)$  umschreiben. Wir nutzen dann die Turmeigenschaft und die Linearität von bedingten Erwartungswerten.
- b. Mit  $C_1 := \sup_x k(x)^{4d+4} < \infty$  gilt  $K_b^4(x - X_i(s)) \leq b^{-4d-4} C_1 \mathbb{1}_{\{|t-t_l|<b\}} \mathbb{1}_{\{|z_j - Z_{ij}(t_l)|<b, j=1, \dots, d\}}$ . Wir nutzen außerdem aus, dass  $\tilde{N}$  und somit  $T_l$  unabhängig von den Daten sind.
- c.  $\mathbb{E} [\mathbb{1}_{\{|z_j - Z_{ij}(t_l)|<b, j=1, \dots, d\}} Y_i(t_l)] = \int_{z-b}^{z+b} \varphi(w) dw \leq (2b)^d \sup_x \varphi(x)$ . Es sei dann  $C_2 := C_1 2^d \sup_x \varphi(x) < \infty$ .
- d. Da  $T_l$  gleichverteilt auf dem Einheitsintervall ist, gilt  $\int \mathbb{1}_{\{|t-t_l|<b\}} dP_{T_l}(t_l) \leq 2b$  (Gleichheit für  $b$  hinreichend klein).  $C_3 := 2C_2$ .

Nun ergibt sich für den unbedingten Erwartungswert

$$\begin{aligned} & \mathbb{E} \left[ \tilde{N}(1)^3 \cdot \int K_b^4(x - X_i(s)) Y_i(s) d\tilde{N}(s) \right] \\ & = \sum_{k=0}^{\infty} P(\tilde{N}(1) = k) \mathbb{E} \left[ \tilde{N}(1)^3 \cdot \int K_b^4(x - X_i(s)) Y_i(s) d\tilde{N}(s) \middle| \tilde{N}(1) = k \right] \\ & = O(b^{-3d-3}) \sum_{k=0}^{\infty} k^4 P(\tilde{N}(1) = k) = O(b^{-3d-3}), \end{aligned}$$

wobei die letzte Identität gilt, da es sich bei der Reihe um das vierte Moment einer Poissonverteilung handelt, welches endlich ist. Damit sind, da die letzte Abschätzung unabhängig von  $x$  ist, auch für  $R_{n1}$  die Voraussetzungen von Behauptung 3 nachgewiesen. Somit ist Behauptung 5 gezeigt und insgesamt auch Lemma 2.4.3.  $\square$

# Kapitel 3

## Semi-Markov-Modelle

In diesem Kapitel wollen wir zeigen, dass sich Semi-Markov-Prozesse derart in den Zählprozess-Modellrahmen von Kapitel 2 einbetten lassen, dass die dort beschriebene und bewiesene Theorie anwendbar ist. Wir erledigen dies zunächst für ein einfaches, progressives Drei-Zustands-Modell jeweils für den Fall ohne und mit Zensierung und anschließend für den allgemeinen Fall zeitheterogener Semi-Markov-Modelle. In allen Fällen besteht der entscheidende Schritt im Nachweis dessen, dass die auf intuitive Art definierte Intensitätsfunktion (multipliziert mit einem Unter-Risiko-Indikator) den Intensitätsprozess eines bestimmten Zählprozesses darstellt. Dies bedeutet insbesondere, dass die Differenz der Prozesse ein Martingal ist.

Eine Bemerkung zur Notation: diese ist in diesem Kapitel separiert von derjenigen des Kapitels 2 zu betrachten. Insbesondere erhalten einige Variablenbezeichnungen hier eine neue Bedeutung.

Um die in diesem Kapitel besprochenen Modelle davon abzugrenzen, sei kurz an die Definition eines Markov-Prozesses erinnert. Anschaulich gesagt ist ein stochastischer Prozess dann markovsch, wenn die Verteilung der Zukunft nur von der Gegenwart abhängt, nicht aber von anderen Aspekten der bisherigen Historie. Formalisiert wird diese Idee durch (siehe Janssen u. Manca, 2006, Definition 3.1)

**Definition 3.0.1** (Markov-Prozess)

*Ein  $\mathbb{S}$ -wertiger stochastischer Prozess  $X$  heißt markovsch, genau dann, wenn für alle  $m \in \mathbb{N}$ ,  $s, s_1, \dots, s_m \in \mathbb{S}$  und alle  $0 \leq t_1 < \dots < t_m < t$  gilt*

$$P(X(t) = s | X(t_k) = s_k, k = 1, \dots, m) = P(X(t) = s | X(t_m) = s_m).$$

Bevor wir auf die einzelnen Modelle eingehen, sollen einige Notationsaspekte erläutert sowie zwei theoretische Lemmata festgehalten werden. Für eine Zufallsvariable  $X$  :

$(\Omega, \mathfrak{A}, P) \rightarrow (\Xi, \mathfrak{B})$  und eine Menge  $A \subset \Omega$  sei  $X|_A : A \rightarrow \Xi$  definiert durch

$$X|_A(\omega) := X(\omega), \quad \omega \in A. \quad (3.1)$$

Dabei sei  $(\Xi, \mathfrak{B})$  ein allgemeiner Messraum.  $X|_A$  heißt *Einschränkung von  $X$  auf  $A$* . Für  $A \in \mathfrak{A}$  sei außerdem

$$\mathfrak{A}|_A := \{B \cap A \mid B \in \mathfrak{A}\} = \mathfrak{A} \cap \mathfrak{P}(A).$$

Dann ist  $\mathfrak{A}|_A$  eine  $\sigma$ -Algebra auf  $A$ . Außerdem ist  $X|_A$  eine  $(\mathfrak{A}|_A, \mathfrak{B})$ -messbare Abbildung. Dementsprechend existiert auch eine von  $X|_A$  erzeugte Teil- $\sigma$ -Algebra  $\sigma(X|_A) \subset \mathfrak{A}|_A$ . Für  $P(A) > 0$  sei weiterhin

$$P_A(B) := P(B), \quad B \in \mathfrak{A}|_A.$$

Dann ist  $P_A$  ein endliches Maß auf  $(A, \mathfrak{A}|_A)$ .

### Lemma 3.0.2

Seien  $X, Y$  Zufallsvariablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{A}, P)$  und  $Z \in L_1(\Omega, \mathfrak{A}, P)$ . Sei außerdem  $A \in \sigma(X) \cap \sigma(Y) \subset \mathfrak{A}$  mit  $P(A) > 0$ . Falls gilt  $\sigma(X|_A) = \sigma(Y|_A)$  ( $\subset \mathfrak{A}|_A$ ), so folgt

$$\mathbb{E}[Z|X]|_A = \mathbb{E}[Z|Y]|_A \quad P_A\text{-f.s.}$$

**Beweis:** Wir zeigen zunächst folgende

*Behauptung (\*)*: Ist  $X : (\Omega, \mathfrak{A}) \rightarrow (\Xi, \mathfrak{B})$  messbar und gilt  $A \in \sigma(X)$ , so folgt

$$\sigma(X|_A) = \sigma(X) \cap \mathfrak{P}(A).$$

Dies bedeutet, dass für  $A \in \sigma(X)$  die von  $X|_A$  erzeugte  $\sigma$ -Algebra gerade aus den Mengen aus  $\sigma(X)$  besteht, welche Teilmengen von  $A$  sind.

*Beweis von (\*)*: „ $\subset$ “: Aus  $C \in \sigma(X|_A)$  folgt die Existenz einer Menge  $B \in \mathfrak{B}$ , sodass

$$C = \{\omega \in A \mid X(\omega) \in B\} = \{\omega \in \Omega \mid X(\omega) \in B\} \cap A.$$

Da beide Mengen des Schnittes in  $\sigma(X)$  liegen, gilt dies auch für  $C$ .

„ $\supset$ “: Aus  $C \in \sigma(X) \cap \mathfrak{P}(A)$  folgt die Existenz einer Menge  $B \in \mathfrak{B}$ , sodass  $C = \{\omega \in \Omega \mid X(\omega) \in B\}$ , sowie, dass  $C \subset A$ . Daraus folgt aber bereits

$$C = \{\omega \in A \mid X(\omega) \in B\} \in \sigma(X|_A),$$

wodurch die Behauptung (\*) bewiesen ist.



Wir wollen nun ausnutzen, dass zwei Funktionen, deren  $\mu$ -Integrale auf allen messbaren Mengen übereinstimmen,  $\mu$ -f.ü. identisch sind (siehe etwa Elstrodt, 2009, Kapitel IV, Satz 4.4). Zum Beweis des Lemmas genügt es also zu zeigen, dass für alle  $B \in \sigma(X|_A)$

$$\int_B \mathbb{E}[Z|X]|_A dP_A = \int_B \mathbb{E}[Z|Y]|_A dP_A.$$

Denn damit wäre die Aussage des Lemmas  $P_A|_{\sigma(X|_A)}$ -f.s. gezeigt und somit auch  $P_A$ -f.s. Wir erhalten

$$\int_B \mathbb{E}[Z|X]|_A dP_A \stackrel{a.}{=} \int_B \mathbb{E}[Z|X] dP \stackrel{b.}{=} \int_B Z dP \stackrel{b.}{=} \int_B \mathbb{E}[Z|Y] dP \stackrel{a.}{=} \int_B \mathbb{E}[Z|Y]|_A dP_A.$$

- a. Für die erste und letzte Identität garantieren  $A \in \mathfrak{A}$ ,  $B \in \mathfrak{A}$  und  $B \subset A$ , dass beide Seiten wohldefiniert sind. Die Gleichheit folgt dann, weil auf seinem Definitionsbereich  $P_A$  gleich  $P$  ist.
- b. Die beiden mittleren Identitäten gelten gemäß der Definition des bedingten Erwartungswertes. Denn nach Behauptung (\*) gilt wegen  $B \in \sigma(X|_A)$  und  $A \in \sigma(X)$  sowohl  $B \in \sigma(X)$  als auch  $B \in \sigma(Y)$ .  $\square$

Um die Voraussetzung der Gleichheit von  $\sigma(X|_A)$  und  $\sigma(Y|_A)$  des Lemmas 3.0.2 später zu überprüfen, benötigen wir das weitere folgende Lemma.

### Lemma 3.0.3

Seien  $\mathcal{S}$  und  $\mathcal{T}$  Indexmengen. Weiter seien  $(\Omega, \mathfrak{A})$ ,  $(\mathcal{X}_s, \mathfrak{B}_s)$ ,  $s \in \mathcal{S}$ , sowie  $(\mathcal{Y}_t, \mathfrak{C}_t)$ ,  $t \in \mathcal{T}$ , messbare Räume. Es seien  $X_s : \Omega \rightarrow \mathcal{X}_s$ ,  $s \in \mathcal{S}$ , und  $Y_t : \Omega \rightarrow \mathcal{Y}_t$ ,  $t \in \mathcal{T}$ , messbare Abbildungen.

Für alle  $s \in \mathcal{S}$  existiere eine Teilmenge  $\mathcal{T}_s \subset \mathcal{T}$  und eine  $(\bigotimes_{t \in \mathcal{T}_s} \mathfrak{C}_t)$ - $\mathfrak{B}_s$ -messbare Abbildung  $g : \times_{t \in \mathcal{T}_s} \mathcal{Y}_t \rightarrow \mathcal{X}_s$  mit

$$X_s = g((Y_t)_{t \in \mathcal{T}_s}).$$

Dann gilt

$$\sigma(X_s, s \in \mathcal{S}) \subset \sigma(Y_t, t \in \mathcal{T}).$$

**Beweis:** Es genügt zu zeigen, dass jedes  $X_s$ ,  $s \in \mathcal{S}$ ,  $\sigma(Y_t, t \in \mathcal{T})$ -messbar ist. Sei dafür  $B \in \mathfrak{B}_s$ . Für dieses gilt

$$X_s^{-1}(B) = \{\omega \in \Omega | X_s(\omega) \in B\} = \{\omega \in \Omega | g((Y_t)_{t \in \mathcal{T}_s}) \in B\} = \{\omega \in \Omega | (Y_t)_{t \in \mathcal{T}_s} \in g^{-1}(B)\}.$$

Es gilt  $g^{-1}(B) \in (\bigotimes_{t \in \mathcal{T}_s} \mathfrak{C}_t)$  wegen der Messbarkeit von  $g$ .  $(Y_t)_{t \in \mathcal{T}_s}$  ist außerdem  $\mathfrak{A}$ - $(\bigotimes_{t \in \mathcal{T}_s} \mathfrak{C}_t)$ -messbar. Damit folgt, dass  $X_s^{-1}(B) \in \sigma(Y_t, t \in \mathcal{T}_s) \subset \sigma(Y_t, t \in \mathcal{T})$ .  $\square$

Neben diesen Lemmata wollen wir außerdem zwei Sätze über Rechenregeln für bedingte Erwartungswerten aus der Literatur zitieren, welche wiederholt angewandt werden.

**Satz 3.0.4** (siehe etwa Witting (1985), Satz 1.122, Satz 1.126)

Seien  $(X, Y)^T : (\Omega, \mathfrak{A}) \rightarrow (\Omega' \times \Omega'', \mathfrak{A}' \otimes \mathfrak{A}'')$ ,  $P_{X|Y=\cdot}(\cdot)$  eine faktorisierte bedingte Verteilung von  $X$  gegeben  $Y$  und

$$g : (\Omega' \times \Omega'', \mathfrak{A}' \otimes \mathfrak{A}'') \rightarrow (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$$

(evtl. uneigentlich)  $P_{(X,Y)}$ -integrabel. Dann gilt

$$\mathbb{E}[g \circ (X, Y) | Y = y] = \int g(x, y) dP_{X|Y=y}(x) \quad \text{für } P_Y\text{-f.a. } y \in \Omega''.$$

Existiert eine bedingte  $\mu$ -Dichte  $f_{X|Y=\cdot}(\cdot)$  von  $X$  gegeben  $Y$ , so gilt außerdem

$$\mathbb{E}[g \circ (X, Y) | Y = y] = \int g(x, y) f_{X|Y=y}(x) d\mu(x) \quad \text{für } P_Y\text{-f.a. } y \in \Omega''.$$

**Satz 3.0.5** (siehe etwa Witting (1985), Satz 1.120)

Seien  $X, \tilde{X} \in \mathcal{L}_1(\Omega, \mathfrak{A}, P)$  und  $Y, \tilde{Y}$  Zufallsgrößen auf  $(\Omega, \mathfrak{A})$ , so dass  $(X, Y)^T$  und  $(\tilde{X}, \tilde{Y})^T$   $P$ -stochastisch unabhängig sind. Dann gilt

$$\mathbb{E}[X\tilde{X} | (Y, \tilde{Y})^T = (\cdot, \cdot)^T] = \mathbb{E}[X | Y = \cdot] \cdot \mathbb{E}[\tilde{X} | \tilde{Y} = \cdot] \quad P_{(Y, \tilde{Y})}\text{-f.s.}$$

### 3.1 Ein einfaches Modell mit drei Zuständen

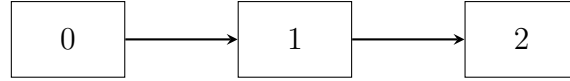
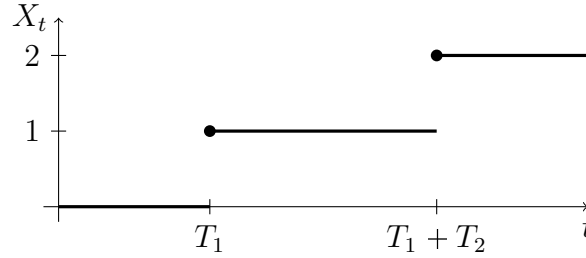
Wir definieren den stochastischen Prozess  $(X_t)_{t \geq 0}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathfrak{F}, P)$  durch

$$X_t := \mathbb{1}_{\{t \geq T_1\}} + \mathbb{1}_{\{t \geq T_1 + T_2\}},$$

wobei  $T_1, T_2$   $P$ -f.s. positive Lebesgue-stetige Zufallsvariablen mit gemeinsamer Dichte  $f_{T_1, T_2}$  seien. Die Marginaldichten seien mit  $f_{T_i}, i = 1, 2$ , bezeichnet, die dazugehörigen Survivalfunktionen mit  $S_{T_i}, i = 1, 2$ . Außerdem seien  $f_{T_2|T_1=t}(d)$  die bedingte Dichte und  $S_{T_2|T_1=t}(d)$  die bedingte Survivalfunktion von  $T_2$  gegeben  $T_1$ .

Die Abbildungen 3.1 und 3.2 veranschaulichen die Struktur von  $X_t$ . Es handelt sich um einen stochastischen Prozess, dessen Pfade rechtsstetig und stückweise konstant sind. Er startet in null, springt in  $T_1$  nach eins und in  $T_1 + T_2$  nach zwei, wobei die Sprungzeitpunkte per Konstruktion f.s. unterschiedlich sind.

Wir wollen die Verteilung des Prozesses wie in der Verweildaueranalyse üblich durch

Abbildung 3.1: Strukturdiagramm des Prozesses  $X_t$ .Abbildung 3.2: Ein exemplarischer Pfad des Prozesses  $X_t$ .

Übergangsintensitäten beschreiben. Diese seien definiert durch

$$\alpha_{01}(t) := \lim_{h \rightarrow 0+} \frac{1}{h} P(X_{t+h} = 1 | X_{t-} = 0), \quad 0 < t,$$

$$\alpha_{12}(t, d) := \lim_{h \rightarrow 0+} \frac{1}{h} P(X_{t+h} = 2 | X_{t-} = 1, T_1 = t - d), \quad 0 < d < t.$$

Während  $\alpha_{01}$  nur eine Funktion der Zeit  $t$  ist, ist  $\alpha_{12}$  eine Funktion in zwei Argumenten: der Zeit  $t$  sowie der Dauer seit Eintritt in Zustand 1  $d$ . An dieser Stelle fließt ein, dass wir keine Markov-Annahme machen. Im Falle eines Markov-Prozesses wäre nämlich auch  $\alpha_{12}$  nur eine Funktion der Zeit und unabhängig von  $d$ .

Wir drücken zunächst diese Übergangsintensitäten unter Verwendung der Dichte- und Survivalfunktionen aus, wobei  $\lambda$  das Lebesgue-Maß bezeichnet:

### Lemma 3.1.1

*Es gilt unter den in diesem Kapitel eingeführten Bezeichnungen:*

(a)  $\alpha_{01}(t) = \frac{f_{T_1}(t)}{S_{T_1}(t)}$  für  $\lambda$ -f.a.  $t$ .

(b) Falls  $f_{T_1, T_2}$  in einer Umgebung von  $(t, 0)$  beschränkt ist, so gilt

$$\alpha_{01}(t) = \lim_{h \rightarrow 0+} \frac{1}{h} P(X_{t+h} = 1 | X_{t-} = 0).$$

(c)  $\alpha_{12}(t, d) = \frac{f_{T_2|T_1=t-d}(d)}{S_{T_2|T_1=t-d}(d)}$  für  $\lambda$ -f.a.  $t, d$ .

**Beweis:** (a) Wir erhalten für  $\lambda$ -f.a.  $t$

$$\begin{aligned} \alpha_{01}(t) &= \lim_{h \rightarrow 0+} \frac{1}{h} P(X_{t+h} = 1 | X_{t-} = 0) = \lim_{h \rightarrow 0+} \frac{1}{h} P(T_1 \leq t + h | T_1 > t) \\ &= \lim_{h \rightarrow 0+} \frac{1}{h} \frac{P(t < T_1 \leq t + h)}{P(T_1 > t)} = \frac{\lim_{h \rightarrow 0+} \frac{1}{h} \int_t^{t+h} f_{T_1}(s) ds}{S_{T_1}(t)} = \frac{f_{T_1}(t)}{S_{T_1}(t)} \end{aligned}$$

(b) Zunächst erhalten wir

$$\begin{aligned} \frac{1}{h}P(X_{t+h} = 1|X_{t-} = 0) &= \frac{1}{h}P(T_1 \leq t+h < T_1 + T_2 | T_1 \geq t) \\ &= \frac{\frac{1}{h}P(t \leq T_1 \leq t+h < T_1 + T_2)}{P(T_1 \geq t)}. \end{aligned}$$

Da der Nenner identisch  $S_{T_1}(t)$  ist, genügt es nach (a) zu zeigen, dass der Zähler für  $h \rightarrow 0+$  gegen  $f_{T_1}(t)$  konvergiert. Dieser Zähler entspricht

$$\begin{aligned} &\frac{1}{h} \int_t^{t+h} \int_{t+h-t_1}^{\infty} f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 \\ &= \frac{1}{h} \int_t^{t+h} \int_0^{\infty} f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 - \frac{1}{h} \int_t^{t+h} \int_0^{t+h-t_1} f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 \\ &= \frac{1}{h} \int_t^{t+h} f_{T_1}(t_1) dt_1 - h^{-1}O(h^2) \\ &\xrightarrow{h \rightarrow 0+} f_{T_1}(t). \end{aligned}$$

Die zweite Gleichheit erklärt sich durch

$$\begin{aligned} 0 &\leq \int_t^{t+h} \int_0^{t+h-t_1} f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 \leq \int_t^{t+h} \int_0^h f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 \\ &\leq h^2 \sup_{(t_1, t_2) \in [t, t+h] \times [0, h]} f_{T_1, T_2}(t_1, t_2) \end{aligned}$$

und die Beschränktheit der gemeinsamen Dichte in einer Umgebung von  $(t, 0)$ .

(c) Wir erhalten für  $\lambda$ -f.a.  $t, d$

$$\begin{aligned} \alpha_{12}(t, d) &= \lim_{h \rightarrow 0+} \frac{1}{h} P(X_{t+h} = 2 | X_{t-} = 1, T_1 = t-d) \\ &= \lim_{h \rightarrow 0+} \frac{1}{h} P(T_1 + T_2 \leq t+h | T_1 < t \leq T_1 + T_2, T_1 = t-d) \\ &= \lim_{h \rightarrow 0+} \frac{1}{h} \frac{P(T_1 < t \leq T_1 + T_2 \leq t+h | T_1 = t-d)}{P(T_1 < t \leq T_1 + T_2 | T_1 = t-d)} \\ &= \lim_{h \rightarrow 0+} \frac{1}{h} \frac{P(t \leq T_1 + T_2 \leq t+h | T_1 = t-d)}{P(t \leq T_1 + T_2 | T_1 = t-d)} \\ &= \lim_{h \rightarrow 0+} \frac{1}{h} \frac{\int \mathbb{1}_{[t, t+h]}(t-d+t_2) f_{T_2|T_1=t-d}(t_2) dt_2}{\int \mathbb{1}_{[t, \infty)}(t-d+t_2) f_{T_2|T_1=t-d}(t_2) dt_2} \\ &= \frac{\lim_{h \rightarrow 0+} \frac{1}{h} \int_d^{d+h} f_{T_2|T_1=t-d}(t_2) dt_2}{\int_d^{\infty} f_{T_2|T_1=t-d}(t_2) dt_2} \\ &= \frac{f_{T_2|T_1=t-d}(d)}{S_{T_2|T_1=t-d}(d)}. \end{aligned}$$

Für die fünfte Identität wurde dabei Satz 3.0.4 mit  $g(x, y) = \mathbb{1}_{[t, \infty)}(x+y)$  genutzt.  $\square$

Im Folgenden wollen wir die Schätzung dieser Intensitäten behandeln, wobei wir uns einzig auf  $\alpha_{12}$  konzentrieren, da nur diese davon betroffen ist, dass wir keine Markov-Annahme machen.

### 3.1.1 Daten

Wir wollen zwei Szenarien möglicher Datenlagen betrachten – zum einen vollständig unzensierte und zum anderen unabhängig, zufällig rechtszensierte Daten.

- **Vollständige Daten:** In diesem Falle nehmen wir an, dass uns  $n$  unabhängig identisch verteilte Realisierungen des Zufallsvektors  $(T_1, T_2)$  zur Verfügung stehen.
- **Zensierte Daten:** Hierfür sei  $U$  eine positive Zufallsvariable, welche stochastisch unabhängig vom Zufallsvektor  $(T_1, T_2)$  ist. Zur Verfügung stehen dann  $n$  unabhängig identisch verteilte Realisierungen des Tripels

$$\begin{aligned} Z_1 &:= T_1 \wedge U, \\ Z_2 &:= (T_1 + T_2) \wedge U \quad \text{und} \\ \Delta &:= \begin{cases} 0, & U < T_1 \\ 1, & T_1 \leq U < T_1 + T_2 \\ 2, & U \geq T_1 + T_2 \end{cases} \end{aligned}$$

### 3.1.2 Zählprozess-Formulierung

Da wir in Kapitel 3.1.3 den Schätzer aus Kapitel 2.1 adaptieren wollen, müssen wir unser Modell in den Zählprozess-Rahmen überführen. Die Definitionen der nötigen Prozesse unterscheiden sich je nach Vorliegen von vollständigen oder zensierten Daten.

- **Vollständige Daten:** Der Prozess

$$Y(t) := Y_1(t) := \mathbb{1}_{\{X_{t-}=1\}} = \mathbb{1}_{\{T_1 < t \leq T_1 + T_2\}}$$

zeigt an, ob der Prozess  $X$  kurz vor  $t$  unter Risiko steht, von 1 nach 2 zu springen.

$$N(t) := N_{12}(t) := \mathbb{1}_{\{T_1 + T_2 \leq t\}}$$

ist ein 0-1-wertiger Zählprozess, der die Anzahl der Sprünge des Prozesses  $X$  von Zustand 1 in Zustand 2 bis zur Zeit  $t$  einschließlich zählt. Schließlich zeigt der Prozess

$$D(t) := (t - T_1) \cdot Y(t)$$

die Dauer an, die der Prozess  $X$  bereits in Zustand 1 verbringt – solange sich der Prozess in Zustand 1 aufhält.

- **Zensierte Daten:** In diesem Falle definieren wir

$$\begin{aligned} Y(t) &:= Y_1(t) := \mathbb{1}_{\{X_{t-}=1, U \geq t\}} = \mathbb{1}_{\{Z_1 < t \leq Z_2\}} = \mathbb{1}_{\{T_1 < t \leq Z_2\}} = \mathbb{1}_{\{T_1 < t \leq T_1 + T_2\} \cap \{U \geq t\}}, \\ N(t) &:= N_{12}(t) := \mathbb{1}_{\{Z_2 \leq t, \Delta=2\}} \quad \text{und} \\ D(t) &:= (t - T_1) \cdot Y(t). \end{aligned}$$

Die Interpretationen sind sehr ähnlich zum unzensierten Fall und müssen nur dahingehend angepasst werden, dass sie sich nur auf beobachtbare Sprünge beziehen. Im Falle, dass Ereignis und Zensierung auf den selben Zeitpunkt fallen, nehmen wir entsprechend der üblichen Konvention an, dass das Ereignis beobachtet wird.

### 3.1.3 Schätzer

In Kapitel 2 haben wir den von NL eingeführten kernbasierten nichtparametrischen Schätzer für Hazardraten unter Berücksichtigung von Kovariaten vorgestellt. Wir wollen diesen hier nutzen, wobei die Verweildauer in Zustand 1 bis zur Zeit  $t$  als eine derartige Kovariable fungiert.

Es sei also  $k$  eine eindimensionale Wahrscheinlichkeitsdichte,  $b > 0$  ein Bandbreitenparameter und  $k_b(\cdot) := b^{-1}k(\cdot/b)$ . Unser Schätzer für die Übergangsintensität  $\alpha_{12}$  lautet damit also in Anlehnung an (2.1)

$$\hat{\alpha}_{12}(t, d) := \frac{\sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) dN_i(s)}{\sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) Y_i(s) ds}, \quad (3.2)$$

wobei die Indizierung in  $i$  die einzelnen unabhängigen Beobachtungen identifiziert, welche bislang nicht in die Notation eingeflossen sind. Die asymptotische Theorie aus den Sätzen 2.1.4, 2.1.5, 2.2.3, 2.2.4 und 2.3.1 können wir auf unser konkretes Modell übertragen. Dafür werden wir nachweisen müssen, dass unser hier präsentiertes Modell in den allgemeineren Modellrahmen von NL passt, wie er in Kapitel 2 vorgestellt wurde. Wir wollen nicht all diese Aussagen für unseren spezifischen Fall wiederholen, aber zumindest die asymptotische Normalität noch einmal explizit formulieren. Die Definition von  $\alpha^*$  aus (2.2) für die Zerlegung (2.3) entspricht hier

$$\alpha_{12}^*(t, d) := \frac{\sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) \alpha_{12}(s, D_i(s)) Y_i(s) ds}{\sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) Y_i(s) ds}.$$

Als Resultat fassen wir zusammen:

**Satz 3.1.2**

Unter den Bedingungen (D), (S) (für  $\alpha(x) = \alpha_{12}(t, d)$  und  $\varphi(x) = \varphi(t, d) = f_{D(t)|Y(t)=1}(d) \cdot P(Y(t) = 1)$ ), sowie (K) und (B) gilt:

$$(a) \quad n^{1/2}b(\hat{\alpha}_{12}(t, d) - \alpha_{12}^*(t, d)) \Rightarrow \mathcal{N}[0, \sigma_{t,d}^2], \text{ wobei } \sigma_{t,d}^2 = \kappa_2^2 \frac{\alpha_{12}(t, d)}{\varphi(t, d)}.$$

$$(b) \quad b^{-2}(\alpha_{12}^*(t, d) - \alpha_{12}(t, d)) \xrightarrow{P} c(t, d), \text{ wobei}$$

$$c(t, d) = \kappa_1 \left[ \frac{(\partial \alpha_{12}(t, d)/\partial t)(\partial \varphi(t, d)/\partial t)}{\varphi(t, d)} + \frac{\partial^2 \alpha_{12}/\partial t^2}{2} + \frac{(\partial \alpha_{12}(t, d)/\partial d)(\partial \varphi(t, d)/\partial d)}{\varphi(t, d)} + \frac{\partial^2 \alpha_{12}/\partial d^2}{2} \right].$$

$$(c) \quad \hat{\sigma}_{t,d}^2 \xrightarrow{P} \sigma_{t,d}^2, \text{ wobei}$$

$$\hat{\sigma}_{t,d}^2 := \left( n^{-1} \sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) Y_i(s) ds \right)^{-2} \cdot n^{-1} b^2 \sum_{i=1}^n \int k_b^2(d - D_i(s)) k_b^2(t - s) dN_i(s).$$

Diese Betrachtungen sind punktwieser Natur und gehen auf Satz 2.1.4 zurück. Aber auch eine multivariate Formulierung (nach Satz 2.2.3), eine gleichmäßige Aussage (nach Satz 2.1.5) oder eine um  $\alpha_{12}$  zentrierte Konvergenzaussage (nach Korollar 2.2.4) sind möglich.

Insbesondere ist es auch möglich, anhand von Kapitel 2.4 statistische Tests der Markov-Eigenschaft des Prozesses  $X_t$  durchzuführen, d.h. daraufhin zu testen, ob die Funktion  $\alpha_{12}(t, d)$  von  $d$  unabhängig ist. Dies gliedert sich in den Rahmen von Kapitel 2.4 ein, indem wir  $\mathcal{B} = \{\beta : [0, 1]^2 \rightarrow \mathbb{R} \mid \beta(t, d_1) = \beta(t, d_2) \ \forall t, d_1, d_2 \in [0, 1]\}$  wählen. Es wird außerdem ein Schätzer  $\hat{\beta}$  benötigt, welcher unter der Nullhypothese  $\alpha_{12} \in \mathcal{B}$  asymptotisch besser ist als  $\hat{\alpha}$ . Dazu können wir einen Schätzer mit der selben Struktur wie (3.2) nutzen, nur ohne die Kern-Terme, die die Kovariate  $D$  berücksichtigen:

$$\hat{\beta}_{12}(t, d) := \hat{\beta}_{12}(t) := \frac{\sum_{i=1}^n \int k_b(t - s) dN_i(s)}{\sum_{i=1}^n \int k_b(t - s) Y_i(s) ds} \quad (3.3)$$

Dies entspricht dem Schätzer, den Hjort (1994) zur nichtparametrischen Schätzung von Hazardraten vorgeschlagen hat, und den NL zur Einbeziehung von Kovariablen verallgemeinert haben. Für die notwendigen asymptotischen Eigenschaften des Schätzers können wir diesen als Spezialfall des Schätzers nach NL mit Dimension des Kovariablenvektors identisch null auffassen.

Für all dies ist die Integration unseres Modells in den Modellrahmen von Kapitel 2

notwendig, was wir im nächsten Abschnitt behandeln werden.

### 3.1.4 Nachweise der Martingaleigenschaft

Wir müssen nachweisen, dass unser Modell in das Framework von NL passt, um dann Satz 2.1.4 (bzw. andere der Resultate) anwenden zu dürfen. Wir wählen als Filtration für beide möglichen Datenlagen  $\mathfrak{F}_t = \sigma(N(u), D(u+), Y(u+), u \leq t)$ . Wie in Kapitel 2.1 angedeutet, weicht dies leicht von der von NL selbst genannten Filtration ab, um deren Rechtsstetigkeit sicherzustellen. Die notwendigen Aussagen fassen wir in einem Satz zusammen:

#### Satz 3.1.3

*Sowohl im Fall zensierter als auch im Fall unzensierter Daten gemäß der Abschnitte 3.1.1 und 3.1.2 ist die Filtration  $\mathfrak{F}_t$  rechtsstetig. Außerdem ist  $\alpha_{12}(t, D(t)) \cdot Y(t)$  vorhersagbar und*

$$M(t) := N(t) - \int_0^t \alpha_{12}(s, D(s))Y(s)ds$$

*ein Martingal, jeweils bezüglich  $\mathfrak{F}_t$ . Dies bedeutet insbesondere, dass  $\alpha_{12}(t, D(t)) \cdot Y(t)$  der Intensitätsprozess von  $N(t)$  ist.*

#### Beweis von Satz 3.1.3 – Vollständige Daten

Zuerst wollen wir die Rechtsstetigkeit der Filtration nachweisen. Laut Theorem 4.2.3 aus Fleming u. Harrington (2011) wäre die Filtration  $\mathfrak{F}_t = \sigma(N(u), D(u+), Y(u+), u \leq t)$  rechtsstetig, falls  $D$  f.s. stückweise konstante Pfade hätte. Wir definieren  $\tilde{D}(t) := T_1 Y(t)$ . Es gilt

$$\mathfrak{F}_t = \sigma(N(u), D(u+), Y(u+), u \leq t) = \sigma(N(u), \tilde{D}(u+), Y(u+), u \leq t)$$

wegen  $\tilde{D}(t) = tY(t) - D(t)$ . Damit wird der genannte Satz anwendbar, weil  $\tilde{D}$  die geforderten stückweise konstanten Pfade besitzt. Die Vorhersagbarkeit von  $\alpha_{12}(t, D(t)) \cdot Y(t)$  folgt aus der Linksstetigkeit von  $Y$  und  $D$  sowie der Stetigkeit von  $\alpha$ . Als nächstes wenden wir uns dem Nachweis der Martingaleigenschaft von  $M$  zu.

Sei  $0 < s < t$ . Zunächst stellen wir fest, dass sich die Zuwächse von  $M$  ergeben durch

$$M(t) - M(s) = \mathbb{1}_{\{s < T_1 + T_2 \leq t\}} - \int_s^t Y(u) \alpha_{12}(u, D(u)) du. \quad (3.4)$$

Die bedingte Erwartung dieser Zuwächse gegeben  $\mathfrak{F}_s$ , von welcher zu zeigen ist, dass sie identisch null ist, werden wir auf drei Mengen separat untersuchen. Diese Mengen sind die Ereignisse, die sich beschreiben lassen durch „Der Prozess ist bis zur Zeit  $s$  bereits zweimal gesprungen.“, „Der Prozess ist bis  $s$  noch gar nicht gesprungen.“, und



„Der Prozess ist bis  $s$  bereits genau einmal gesprungen.“:

$$\begin{aligned} \mathbb{E}[M(t) - M(s)|\mathfrak{F}_s] &= \mathbb{1}_{\{T_1+T_2 \leq s\}} \mathbb{E}[M(t) - M(s)|\mathfrak{F}_s] \Big|_{\{T_1+T_2 \leq s\}} \\ &\quad + \mathbb{1}_{\{T_1 > s\}} \mathbb{E}[M(t) - M(s)|\mathfrak{F}_s] \Big|_{\{T_1 > s\}} \\ &\quad + \mathbb{1}_{\{T_1 \leq s < T_1+T_2\}} \mathbb{E}[M(t) - M(s)|\mathfrak{F}_s] \Big|_{\{T_1 \leq s < T_1+T_2\}} \end{aligned} \quad (3.5)$$

Es sei hier an die Definition aus Formel (3.1) erinnert. Auf der Menge  $\{T_1 + T_2 \leq s\}$  ist offensichtlich  $\mathbb{1}_{\{s < T_1+T_2 \leq t\}} = 0$ . Ebenfalls gilt auf dieser Menge  $Y(u) = 0$ . Damit folgt durch Einsetzen in (3.4)

$$\mathbb{1}_{\{T_1+T_2 \leq s\}} \mathbb{E}[M(t) - M(s)|\mathfrak{F}_s] \Big|_{\{T_1+T_2 \leq s\}} = 0.$$

Auf der Menge  $\{T_1 > s\}$  müssen die bedingten Erwartungen konstant im Argument  $\omega$  sein. Dies ist etwa einzusehen, indem man bemerkt, dass sich nach dem Faktorisierungslemma (siehe etwa Bauer, 1992, S. 71, Satz 11.7) der bedingte Erwartungswert als Verkettung einer messbaren Funktion mit  $(N(u), D(u+), Y(u+))_{u \leq s}$  darstellen lässt. Diese Zufallsvariablen sind aber für alle  $\omega \in \{T_1 > s\}$  identisch null.

Sei also zunächst  $\mathbb{E}[\mathbb{1}_{\{s < T_1+T_2 \leq t\}}|\mathfrak{F}_s] \Big|_{\{T_1 > s\}} =: c$ . Für dieses  $c$  gilt gemäß der Definition bedingter Erwartungswerte

$$\begin{aligned} c \cdot P(T_1 > s) &= \int_{\{T_1 > s\}} c dP = \int_{\{T_1 > s\}} \mathbb{1}_{\{s < T_1+T_2 \leq t\}} dP \\ &= P(s < T_1, T_1 + T_2 \leq t). \end{aligned}$$

Weiter gilt für  $\mathbb{E} \left[ \int_s^t Y(u) \alpha_{12}(u, D(u)) du | \mathfrak{F}_s \right] \Big|_{\{T_1 > s\}} =: \tilde{c}$

$$\begin{aligned} \tilde{c} \cdot P(T_1 > s) &= \int_{\{T_1 > s\}} \tilde{c} dP = \int_{\{T_1 > s\}} \int_s^t Y(u) \alpha_{12}(u, D(u)) du dP \\ &\stackrel{a.}{=} \mathbb{E}_{T_1} \left[ \mathbb{E} \left[ \int_s^t \mathbb{1}_{\{s < T_1 < u \leq T_1+T_2\}} \alpha_{12}(u, u - T_1) du \Big| T_1 = t_1 \right] \right] \\ &\stackrel{b.}{=} \mathbb{E}_{T_1} \left[ \int_s^t \mathbb{1}_{(s,u)}(t_1) \alpha_{12}(u, u - t_1) P(T_1 + T_2 \geq u | T_1 = t_1) du \right] \\ &\stackrel{c.}{=} \mathbb{E}_{T_1} \left[ \int_s^t \mathbb{1}_{(s,u)}(t_1) \frac{f_{T_2|T_1=t_1}(u - t_1)}{S_{T_2|T_1=t_1}(u - t_1)} S_{T_2|T_1=t_1}(u - t_1) du \right] \\ &\stackrel{d.}{=} \mathbb{E}_{T_1} \left[ \mathbb{1}_{(s,\infty)}(t_1) \int_{t_1}^t f_{T_2|T_1=t_1}(u - t_1) du \right] \\ &\stackrel{e.}{=} \mathbb{E}_{T_1} [\mathbb{1}_{(s,\infty)}(t_1) P(T_1 + T_2 \leq t | T_1 = t_1)] \\ &\stackrel{f.}{=} P(s < T_1, T_1 + T_2 \leq t), \end{aligned}$$

wobei hier die Notation  $\mathbb{E}_{T_1}[\dots] = \int \dots dP_{T_1}(t_1)$  Verwendung findet.

- a. Die erste Integrationsmenge wird zu einer Indikatorfunktion umgeschrieben, welche mit  $Y(u)$  zusammengefasst wird. Dadurch können wir das äußere Integral als Erwartungswert schreiben und gemäß des Iterationssatzes für bedingte Erwartungswerte zunächst eine bedingte Erwartung bilden. Das Argument  $D(u)$  können wir durch  $u - T_1$  ersetzen, weil beide auf  $\{Y(u) \neq 0\}$  identisch sind.
- b. Zunächst ziehen wir den bedingten Erwartungswert in das innere Integral, dann nutzen wir den Multiplikationssatz für bedingte Erwartungswerte, um  $T_1$ -messbare Faktoren herauszuziehen. Der übrige bedingte Erwartungswert geht in eine bedingte Wahrscheinlichkeit über.
- c. Gemäß Lemma 3.1.1 sowie der Rechnung

$$\begin{aligned} P(T_1 + T_2 \geq u | T_1 = t_1) &= \int \mathbb{1}_{[u-t_1, \infty)}(t_2) f_{T_2|T_1=t_1}(t_2) dt_2 \\ &= \int_{u-t_1}^{\infty} f_{T_2|T_1=t_1}(t_2) dt_2 = S_{T_2|T_1=t_1}(u - t_1), \end{aligned} \quad (3.6)$$

welche Satz 3.0.4 für die erste Identität nutzt.

- d. Die bedingten Survival-Funktionen werden gekürzt. Mit  $\mathbb{1}_{(s,u)}(t_1) = \mathbb{1}_{(s,\infty)}(t_1) \mathbb{1}_{(t_1,\infty)}(u)$  wird die zweite dieser Indikatorfunktionen in die Integralgrenze übertragen.
- e. Dies geht auf folgende Rechnung zurück, die der aus Schritt c. ähnelt:

$$\begin{aligned} P(T_1 + T_2 \leq t | T_1 = t_1) &= \int \mathbb{1}_{[0,t]}(t_1 + t_2) f_{T_2|T_1=t_1}(t_2) dt_2 \\ &= \int_0^{t-t_1} f_{T_2|T_1=t_1}(t_2) dt_2 = \int_{t_1}^t f_{T_2|T_1=t_1}(u - t_1) du \end{aligned}$$

Im letzten Schritt wurde hier  $t_2 = u - t_1$  substituiert.

- f. Der letzte Schritt ergibt sich mit folgenden Zwischenschritten:

$$\begin{aligned} \mathbb{E}_{T_1} [\mathbb{1}_{(s,\infty)}(t_1) P(T_1 + T_2 \leq t | T_1 = t_1)] &= \mathbb{E} [\mathbb{1}_{\{T_1 > s\}} P(T_1 + T_2 \leq t | T_1)] \\ &= \mathbb{E} [\mathbb{E} [\mathbb{1}_{\{T_1 > s, T_1 + T_2 \leq t\}} | T_1]] = P(s < T_1, T_1 + T_2 \leq t). \end{aligned}$$

Damit ist gezeigt, dass  $c = \tilde{c}$  und somit

$$\mathbb{1}_{\{T_1 > s\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{T_1 > s\}} = 0.$$

Zu untersuchen bleibt die bedingte Erwartung auf der Menge  $\{T_1 \leq s < T_1 + T_2\}$ . Wir wollen zunächst feststellen, dass

$$\mathbb{E}[\cdot | \mathfrak{F}_s] \Big|_{\{T_1 \leq s < T_1 + T_2\}} = \mathbb{E}[\cdot | T_1, \mathbb{1}_{\{T_1 + T_2 > s\}}] \Big|_{\{T_1 \leq s < T_1 + T_2\}}.$$

Die heuristische Idee ist folgende: Wenn wir wissen, dass das Ereignis  $\{T_1 \leq s < T_1 + T_2\}$  eingetreten ist, ist die einzige weitere Information, die wir zur Rekonstruktion des Prozesses bis zur Zeit  $s$  benötigen, die Ausprägung von  $T_1$ . Die Indikatorfunktion muss der Bedingung hinzugefügt werden, damit die Menge  $\{T_1 \leq s < T_1 + T_2\}$  messbar bezüglich der  $\sigma$ -Algebra der Bedingung ist. Formalisiert wird diese Idee durch Lemma 3.0.2.

Um zu sehen, dass die Voraussetzungen des Lemmas erfüllt sind, sei zunächst an die Definition  $\mathfrak{F}_s = \sigma(N(u), D(u+), Y(u+), u \leq s)$  erinnert. Zuerst gilt sowohl  $\{T_1 \leq s < T_1 + T_2\} = \{Y(s) = 1\} \in \mathfrak{F}_s$  als auch  $\{T_1 \leq s < T_1 + T_2\} = \{T_1 \leq s, \mathbb{1}_{\{T_1 + T_2 > s\}} = 1\} \in \sigma(T_1, \mathbb{1}_{\{T_1 + T_2 > s\}})$ . Für die Anwendung von Lemma 3.0.2 ist also zu prüfen, dass  $\sigma(T_1, \mathbb{1}_{\{T_1 + T_2 > s\}} |_{\{T_1 \leq s < T_1 + T_2\}}) = \sigma((N(u), D(u+), Y(u+), u \leq s) |_{\{T_1 \leq s < T_1 + T_2\}})$ . Für diese Prüfung nutzen wir Lemma 3.0.3. Wir definieren die  $Y(u)$ -messbaren Zufallsvariablen  $W_u := u \cdot \mathbb{1}_{\{Y(u)=1\}} + \infty \cdot \mathbb{1}_{\{Y(u)=0\}}$ . Für alle  $\omega \in \{T_1 \leq s < T_1 + T_2\}$  gilt einerseits

$$T_1(\omega) = \inf\{u \leq s | Y(u, \omega) = 1\} = \inf\{W(u, \omega) | u \leq s\} = \inf\{W(q, \omega) | q \leq s, q \in \mathbb{Q}\},$$

$$\mathbb{1}_{\{T_1 + T_2 > s\}}(\omega) \equiv 1$$

und andererseits

$$\begin{aligned} N(u, \omega) &= 0, \quad u \leq s, \\ Y(u, \omega) &= \mathbb{1}_{\{T_1(\omega) \leq u\}}, \quad u \leq s, \\ Y(u+, \omega) &= Y(u, \omega) \\ D(u, \omega) &= (u - T_1(\omega)) \cdot Y(u, \omega), \quad u \leq s. \\ D(u+, \omega) &= D(u, \omega) \end{aligned}$$

Da insbesondere  $\inf : \bar{\mathbb{R}}^{\mathbb{N}} \rightarrow \bar{\mathbb{R}}$  Borel-messbar ist (siehe etwa Chung, 2001, Theorem 3.1.6), folgt daraus, dass Lemma 3.0.3 anwendbar ist. Wir betrachten nun zunächst

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{\{s < T_1 + T_2 \leq t\}} | T_1 = t_1] &= \int \mathbb{1}_{(s, t]}(t_1 + u) f_{T_2 | T_1 = t_1}(u) du \\ &= \int \mathbb{1}_{(s-t_1, t-t_1]}(u) f_{T_2 | T_1 = t_1}(u) du = \int_{s-t_1}^{t-t_1} f_{T_2 | T_1 = t_1}(u) du \\ &= \int_{(s-t_1) \vee 0}^{t-t_1} f_{T_2 | T_1 = t_1}(u) du = \int_{s \vee t_1}^t f_{T_2 | T_1 = t_1}(u - t_1) du, \end{aligned}$$

wobei wir erneut Satz 3.0.4 ausnutzen sowie die (fast sichere) Positivität von  $T_2$ .

Außerdem gilt

$$\mathbb{E}[\mathbb{1}_{\{s < T_1 + T_2 \leq t\}} | T_1 = t_1, \mathbb{1}_{\{T_1 + T_2 > s\}} = 1] = \frac{\mathbb{E}[\mathbb{1}_{\{s < T_1 + T_2 \leq t\}} | T_1 = t_1]}{P(T_1 + T_2 > s | T_1 = t_1)},$$

da  $\{s < T_1 + T_2 \leq t\} \subset \{T_1 + T_2 > s\}$ . Die allgemeine Idee hier lautet, dass für  $A \subset B$

$$\mathbb{E}[\mathbb{1}_A | T = t, \mathbb{1}_B = 1] = P(A | T = t, B) = \frac{P(A \cap B | T = t)}{P(B | T = t)} = \frac{\mathbb{E}[\mathbb{1}_A | T = t]}{P(B | T = t)}.$$

Bedingt auf  $\mathbb{1}_{\{T_1 + T_2 > s\}} = 0$  ist das Ergebnis gemäß analoger Argumentation identisch null. Weiter erhalten wir ebenfalls

$$\begin{aligned} \mathbb{E} \left[ \int_s^t Y(u) \alpha_{12}(u, D(u)) du \middle| T_1 = t_1 \right] &\stackrel{a.}{=} \int_s^t \alpha_{12}(u, u - t_1) \mathbb{E}[\mathbb{1}_{\{T_1 < u \leq T_1 + T_2\}} | T_1 = t_1] du \\ &\stackrel{b.}{=} \int_s^t \mathbb{1}_{(t_1, \infty)}(u) \frac{f_{T_2 | T_1 = t_1}(u - t_1)}{S_{T_2 | T_1 = t_1}(u - t_1)} P(u \leq T_1 + T_2 | T_1 = t_1) du \\ &\stackrel{c.}{=} \int_{s \vee t_1}^t \frac{f_{T_2 | T_1 = t_1}(u - t_1)}{S_{T_2 | T_1 = t_1}(u - t_1)} S_{T_2 | T_1 = t_1}(u - t_1) du \\ &= \int_{s \vee t_1}^t f_{T_2 | T_1 = t_1}(u - t_1) du \end{aligned}$$

- a. Wir vertauschen die Integrationsreihenfolge und ziehen den  $T_1$ -messbaren Faktor  $\alpha_{12}$  aus dem bedingten Erwartungswert.
- b. Erneutes Herausziehen eines  $T_1$ -messbaren Faktors sowie Lemma 3.1.1.
- c. Siehe (3.6).

Auch hier gilt

$$\mathbb{E} \left[ \int_s^t Y(u) \alpha_{12}(u, D(u)) du \middle| T_1 = t_1, \mathbb{1}_{\{T_1 + T_2 > s\}} = 1 \right] = \frac{\mathbb{E} \left[ \int_s^t Y(u) \alpha_{12}(u, D(u)) du \middle| T_1 = t_1 \right]}{P(T_1 + T_2 > s | T_1 = t_1)},$$

zu sehen am besten in der Formel nach Schritt a. in der vorigen Gleichungskette, da  $\{T_1 < u \leq T_1 + T_2\} \subset \{T_1 + T_2 > s\}$  für alle  $u > s$ . Damit folgt insgesamt

$$\mathbb{E}[M(t) - M(s) | T_1 = \cdot, \mathbb{1}_{\{T_1 + T_2 > s\}} = \cdot] \Big|_{\{T_1 \leq s < T_1 + T_2\}} = 0$$

und daher auch

$$\mathbb{1}_{\{T_1 \leq s < T_1 + T_2\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{T_1 \leq s < T_1 + T_2\}} = 0.$$

Dies schließt den Beweis der Martingaleigenschaft von  $M(t)$ .

**Beweis von Satz 3.1.3 – Zensierte Daten**

Die Nachweise der Rechtsstetigkeit von  $\mathfrak{F}_t$  und der Vorhersagbarkeit von  $\alpha Y$  sind identisch zum unzensierten Fall. Auch die Argumentation für die Martingaleigenschaft verläuft ähnlich. Wieder sei  $0 < s < t$ . Analog zu (3.4) stellen wir fest, dass

$$M(t) - M(s) = \mathbb{1}_{\{s < Z_2 \leq t, \Delta=2\}} - \int_s^t Y(u) \alpha_{12}(u, D(u)) du. \quad (3.7)$$

Analog zu (3.5) untersuchen wir die bedingte Erwartung des Zuwachses auf drei disjunkten Ereignissen separat:

$$\begin{aligned} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] &= \mathbb{1}_{\{Z_2 \leq s\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{Z_2 \leq s\}} \\ &\quad + \mathbb{1}_{\{Z_1 > s\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{Z_1 > s\}} \\ &\quad + \mathbb{1}_{\{Z_1 \leq s < Z_2\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{Z_1 \leq s < Z_2\}} \end{aligned}$$

Auf der Menge  $\{Z_2 \leq s\}$  gilt offensichtlich  $\mathbb{1}_{\{s < Z_2 \leq t, \Delta=2\}} = 0$ . Da auf diesem Ereignis  $Y(t) \equiv 0$  gilt, folgt mit (3.7) direkt

$$\mathbb{1}_{\{Z_2 \leq s\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{Z_2 \leq s\}} = 0.$$

Auf dem Ereignis  $\{Z_1 > s\}$  müssen, mit dem gleichen Argument wie im unzensierten Fall, bedingte Erwartungen konstant sein. Für  $c := \mathbb{E}[\mathbb{1}_{\{s < Z_2 \leq t, \Delta=2\}} | \mathfrak{F}_s] \Big|_{\{Z_1 > s\}}$  erhalten wir gemäß der Definition bedingter Erwartungswerte

$$\begin{aligned} c \cdot P(Z_1 > s) &= \int_{\{Z_1 > s\}} c dP = \int_{\{Z_1 > s\}} \mathbb{1}_{\{s < Z_2 \leq t, \Delta=2\}} dP \\ &= P(s < Z_1 < Z_2 \leq t, T_1 + T_2 \leq U). \end{aligned}$$

Des Weiteren gilt für  $\tilde{c} := \mathbb{E} \left[ \int_s^t Y(v) \alpha_{12}(v, D(v)) dv \middle| \mathfrak{F}_s \right] \Big|_{\{Z_1 > s\}}$

$$\begin{aligned}
\tilde{c} \cdot P(Z_1 > s) &= \int_{\{Z_1 > s\}} \tilde{c} dP = \int_{\{Z_1 > s\}} \int_s^t Y(v) \alpha_{12}(v, D(v)) dv dP \\
&\stackrel{a.}{=} \mathbb{E}_{T_1} \left[ \mathbb{E} \left[ \int_s^t \mathbb{1}_{\{s < T_1 < v \leq Z_2\}} \alpha_{12}(v, v - T_1) dv \middle| T_1 = t_1 \right] \right] \\
&\stackrel{b.}{=} \mathbb{E}_{T_1} \left[ \int_s^t \alpha(v, v - t_1) P \left( (T_1 + T_2) \wedge U \geq v \middle| T_1 = t_1 \right) \mathbb{1}_{(s,v)}(t_1) dv \right] \\
&\stackrel{c.}{=} \mathbb{E}_{T_1} \left[ \int_s^t \frac{f_{T_2|T_1=t_1}(v - t_1)}{S_{T_2|T_1=t_1}(v - t_1)} S_{T_2|T_1=t_1}(v - t_1) P(U \geq v | T_1 = t_1) \mathbb{1}_{(s,v)}(t_1) dv \right] \\
&\stackrel{d.}{=} \mathbb{E}_{T_1} \left[ \mathbb{1}_{(s,\infty)}(t_1) \int_{t_1}^t f_{T_2|T_1}(v - t_1) P(U \geq v | T_1 = t_1) dv \right] \\
&\stackrel{e.}{=} \mathbb{E}_{T_1} \left[ \mathbb{1}_{(s,\infty)}(t_1) P(T_1 + T_2 \leq t, U \geq T_1 + T_2 | T_1 = t_1) \right] \\
&\stackrel{f.}{=} P(s < T_1 < T_1 + T_2 \leq t, U \geq T_1 + T_2) \\
&= P(s < Z_1 < Z_2 \leq t, T_1 + T_2 \leq U).
\end{aligned}$$

- $Y(v)$  wird mit dem Integrationsbereich zu einer einzelnen Indikatorfunktion umgeschrieben. Das äußere Integral ist nun ein Erwartungswert, wobei wir noch eine bedingte Erwartung einfügen. Außerdem gilt  $D(v) = v - T_1$  auf dem Ereignis  $\{Y(v) \neq 0\}$ .
- Wir vertauschen zunächst die Integrationsreihenfolge und nutzen dann den Multiplikationssatz für bedingte Erwartungswerte um die  $T_1$ -messbare Faktoren aus dem bedingten Erwartungswert zu ziehen.
- Wir nutzen Lemma 3.1.1 sowie die Identität

$$\begin{aligned}
P \left( (T_1 + T_2) \wedge U \geq v \middle| T_1 = t_1 \right) &= P \left( T_1 + T_2 \geq v, U \geq v \middle| T_1 = t_1 \right) \\
&= P \left( T_1 + T_2 \geq v \middle| T_1 = t_1 \right) P(U \geq v) \\
&= S_{T_2|T_1=t_1}(v - t_1) P(U \geq v | T_1 = t_1).
\end{aligned} \tag{3.8}$$

Die Faktorisierung folgt gemäß Satz 3.0.5, da die Zufallsvektoren

$$\begin{pmatrix} \mathbb{1}_{\{T_1 + T_2 \geq v\}} \\ T_1 \end{pmatrix} \text{ und } \begin{pmatrix} \mathbb{1}_{\{U \geq v\}} \\ 1 \end{pmatrix}$$

annahmegemäß stochastisch unabhängig sind. Im Anschluss muss wieder Rechnung (3.6) angewandt werden.

- Kürzen, sowie  $\mathbb{1}_{(s,v)}(t_1) = \mathbb{1}_{(s,\infty)}(t_1) \mathbb{1}_{(t_1,\infty)}(v)$ . Der zweite Indikator wird in die Integralgrenze gezogen.

- e. Wir stellen zunächst fest, dass  $T_2$  und  $U$  auch bedingt auf  $\{T_1 = t_1\}$  stochastisch unabhängig sind. Dies folgt analog zur Faktorisierung in Schritt c. dieser Aufzählung (gemäß Satz 3.0.5). Damit erhalten wir

$$\begin{aligned}
& P(T_1 + T_2 \leq t, U \geq T_1 + T_2 | T_1 = t_1) \\
&= \int \int \mathbb{1}_{(0,t]}(t_1 + t_2) \mathbb{1}_{[t_1+t_2,\infty)}(u) dP_{(U,T_2)|T_1=t_1}(u, t_2) \\
&= \int \mathbb{1}_{(-t_1, t-t_1]}(t_2) \left[ \int \mathbb{1}_{[t_1+t_2,\infty)}(u) dP_{U|T_1=t_1}(u) \right] dP_{T_2|T_1=t_1}(t_2) \\
&= \int_0^{t-t_1} P(U \geq t_1 + t_2 | T_1 = t_1) f_{T_2|T_1=t_1}(t_2) dt_2 \\
&= \int_{t_1}^t P(U \geq v | T_1 = t_1) f_{T_2|T_1=t_1}(v - t_1) dv.
\end{aligned}$$

Zunächst wird dabei wieder Satz 3.0.4 angewandt. Die dritte Identität nutzt aus, dass  $T_2$  eine strikt positive Zufallsvariable ist.

- f. Die Berechnung des äußeren Erwartungswertes erklärt sich mit den Zwischenschritten

$$\begin{aligned}
& \mathbb{E}_{T_1} [\mathbb{1}_{(s,\infty)}(t_1) P(T_1 + T_2 \leq t, U \geq T_1 + T_2 | T_1 = t_1)] \\
&= \mathbb{E} [\mathbb{1}_{\{T_1 > s\}} \mathbb{E} (\mathbb{1}_{\{T_1+T_2 \leq t, U \geq T_1+T_2\}} | T_1)] = \mathbb{E} [\mathbb{E} (\mathbb{1}_{\{T_1 > s, T_1+T_2 \leq t, U \geq T_1+T_2\}} | T_1)] \\
&= P(s < T_1 < T_1 + T_2 \leq t, U \geq T_1 + T_2)
\end{aligned}$$

Zu untersuchen bleibt die bedingte Erwartung auf der Menge  $\{Z_1 \leq s < Z_2\}$ . Aus zum unzensierten Fall analogen Gründen gilt hier

$$\mathbb{E}[\cdot | \mathfrak{F}_s] \big|_{\{Z_1 \leq s < Z_2\}} = \mathbb{E}[\cdot | T_1, \mathbb{1}_{\{Z_2 > s\}}] \big|_{\{Z_1 \leq s < Z_2\}}.$$

Wir betrachten nun zunächst

$$\begin{aligned}
& \mathbb{E}[\mathbb{1}_{\{s < Z_2 \leq t, \Delta=2\}} | T_1 = t_1] = \int \int \mathbb{1}_{(s,t]}(t_1 + t_2) \mathbb{1}_{[t_1+t_2,\infty)}(u) dP_{(U,T_2)|T_1=t_1}(u, t_2) \\
&= \int \mathbb{1}_{(s-t_1, t-t_1]}(t_2) \left[ \int \mathbb{1}_{[t_1+t_2,\infty)}(u) dP_{U|T_1=t_1}(u) \right] dP_{T_2|T_1=t_1}(t_2) \\
&= \int_{s-t_1}^{t-t_1} P(U \geq t_1 + t_2 | T_1 = t_1) f_{T_2|T_1=t_1}(t_2) dt_2 \\
&= \int_{s \vee t_1}^t P(U \geq v | T_1 = t_1) f_{T_2|T_1=t_1}(v - t_1) dv
\end{aligned}$$

wobei wir erneut Satz 3.0.4 ausnutzen sowie die Tatsache, dass  $T_2$  fast sicher positiv ist.

Wir erhalten ebenfalls

$$\begin{aligned}
\mathbb{E} \left[ \int_s^t Y(v) \alpha_{12}(v, D(v)) dv \middle| T_1 = t_1 \right] &\stackrel{a.}{=} \int_s^t \alpha_{12}(v, v - t_1) \mathbb{E}[\mathbf{1}_{\{T_1 < v \leq Z_2\}} | T_1 = t_1] dv \\
&\stackrel{b.}{=} \int_s^t \mathbf{1}_{(t_1, \infty)}(v) \frac{f_{T_2|T_1=t_1}(v - t_1)}{S_{T_2|T_1=t_1}(v - t_1)} P((T_1 + T_2) \wedge U \geq v | T_1 = t_1) dv \\
&\stackrel{c.}{=} \int_{s \vee t_1}^t \frac{f_{T_2|T_1=t_1}(v - t_1)}{S_{T_2|T_1=t_1}(v - t_1)} S_{T_2|T_1=t_1}(v - t_1) P(U \geq v | T_1 = t_1) dv \\
&= \int_{s \vee t_1}^t f_{T_2|T_1=t_1}(v - t_1) P(U \geq v | T_1 = t_1) dv.
\end{aligned}$$

- a. Wir vertauschen die Integrationsreihenfolge und ziehen den  $T_1$ -messbaren Faktor  $\alpha_{12}$  aus dem bedingten Erwartungswert.
- b. Erneutes Herausziehen eines  $T_1$ -messbaren Faktors sowie Lemma 3.1.1.
- c. Siehe (3.8) in Verbindung mit (3.6).

Aus zum unzensierten Fall analogen Gründen gilt hier

$$\mathbb{E}[\mathbf{1}_{\{s < Z_2 \leq t, \Delta=2\}} | T_1 = t_1, \mathbf{1}_{\{Z_2 > s\}} = 1] = \frac{\mathbb{E}[\mathbf{1}_{\{s < Z_2 \leq t, \Delta=2\}} | T_1 = t_1]}{P(Z_2 > s | T_1 = t_1)}$$

und

$$\mathbb{E} \left[ \int_s^t Y(v) \alpha_{12}(v, D(v)) dv \middle| T_1 = t_1, \mathbf{1}_{\{Z_2 > s\}} = 1 \right] = \frac{\mathbb{E} \left[ \int_s^t Y(v) \alpha_{12}(v, D(v)) dv \middle| T_1 = t_1 \right]}{P(Z_2 > s | T_1 = t_1)}.$$

Damit folgt insgesamt

$$\mathbb{E}[M(t) - M(s) | T_1 = \cdot, \mathbf{1}_{\{Z_2 > s\}} = \cdot] \Big|_{\{Z_1 \leq s < Z_2\}} = 0$$

und daher auch

$$\mathbf{1}_{\{Z_1 \leq s < Z_2\}} \mathbb{E}[M(t) - M(s) | \mathfrak{F}_s] \Big|_{\{Z_1 \leq s < Z_2\}} = 0.$$

Dies vollendet den Nachweis der Martingaleigenschaft von  $M(t)$ .

## 3.2 Allgemeine Semi-Markov-Prozesse

Semi-Markov-Prozesse sind Mehrzustandsmodelle, bei denen die zukünftige Verteilung vom aktuellen Zustand und von der Verweildauer im aktuellen Zustand abhängen darf, nicht aber von anderen Aspekten der Historie des Prozesses. Wir betrachten heterogene Prozesse, bei denen außerdem Übergangsintensitäten mit der (deterministischen) Zeit variieren. Siehe etwa Devolder u. a. (2015), Kapitel 4, für einen Überblick, oder Janssen



u. Manca (2006) für weitere Details. Siehe Keiding (1986) für einen frühen Zählprozess-Ansatz in diesem Kontext.

Es sei zunächst  $\mathbb{S}$  der endliche Zustandsraum des Prozesses und  $(\Omega, \mathfrak{F}, P)$  ein Wahrscheinlichkeitsraum. Wir definieren einige Zufallsvariablen.

- $T_1, T_2, \dots$  sei eine Folge  $P$ -f.s. positiver Zufallsvariablen.
- $Z_m := \sum_{j=1}^m T_j, m \in \mathbb{N}_0$ , insbesondere  $Z_0 = 0$ .
- $S_0, S_1, \dots$ , sei eine Folge  $\mathbb{S}$ -wertiger Zufallsvariablen mit  $S_j \neq S_{j+1}, j \in \mathbb{N}_0$ .

Nun definieren wir den Prozess  $X$  durch

$$X(t) := \sum_{m=0}^{\infty} \mathbf{1}_{[Z_m, Z_{m+1})}(t) S_m, \quad t > 0.$$

Der Prozess  $X$  springt also zu den Zeitpunkten  $Z_m$  zwischen den endlich vielen Zuständen der Menge  $\mathbb{S}$ . Auf diese Art und Weise kann ein allgemeines Mehrzustandsmodell definiert werden. Die oben skizzierte Semi-Markov-Eigenschaft wird durch folgende Definition formalisiert.

**Definition 3.2.1**

*Der Prozess  $X$  ist ein (heterogener) Semi-Markov-Prozess genau dann, wenn für alle  $\{r \neq s\} \subset \mathbb{S}$  und alle  $x, z > 0$  gilt*

$$\begin{aligned} P(S_{m+1} = s, T_{m+1} \leq x | (S_0, Z_0), \dots, (S_m, Z_m) = (r, z)) \\ &= P(S_{m+1} = s, T_{m+1} \leq x | S_m = r, Z_m = z) \\ &= P(S_2 = s, T_2 \leq x | S_1 = r, Z_1 = z) \\ &=: F_z^{rs}(x). \end{aligned}$$

Die Definition hat zwei entscheidende Teile. Zum einen die Unabhängigkeit der Zukunft von den „nicht neuesten“ Informationen  $(S_0, Z_0), \dots, (S_{m-1}, Z_{m-1})$ , zum anderen die Unabhängigkeit der Übergangswahrscheinlichkeiten von  $m$ . Wir führen für  $r \in \mathbb{S}$  einige zusätzliche Bezeichnungen ein.

- $F_z^{rr} := 0$
- $F_z^r(x) := \sum_{s \in \mathbb{S}} F_z^{rs}(x)$ . Dies ist die bedingte Verteilungsfunktion von  $T_m$ , gegeben die Vergangenheit.
- $S_z^r(x) := 1 - F_z^r(x)$ . Dies ist die dazugehörige Survival-Funktion.

Wir werden außerdem im Folgenden annehmen, dass die oben genannten Funktionen differenzierbar sind und definieren für  $\{r \neq s\} \subset \mathbb{S}, z > 0$

$$f_z^{rs} := (F_z^{rs})'.$$

Darüber hinaus formulieren wir eine zusätzliche Bedingung, die sich an die später als Übergangsintensitäten identifizierten Terme richtet.

(L)  $f_z^{rs}(x)/S_z^r(x)$  sind gleichmäßig beschränkt über  $r, s \in \mathbb{S}$ ,  $z, x \in [0, 1]$ .

Es sei darauf hingewiesen, dass damit wegen  $S_z^r(x) \leq 1$  gleichzeitig auch  $f_z^{rs}(x)$  gleichmäßig beschränkt sind. Wir definieren zwei Zufallsvariablen, um uns der Definition von Übergangsintensitäten zu nähern.

- $\mu(t) := \sum_{m=0}^{\infty} m \cdot \mathbb{1}_{[Z_m, Z_{m+1})}(t) = \sum_{m=1}^{\infty} \mathbb{1}_{\{Z_m \geq t\}}.$
- $\tilde{D}(t) := t - \sup(s \leq t | X(s) \neq X(t)) = \sum_{m=0}^{\infty} (t - Z_m) \mathbb{1}_{[Z_m, Z_{m+1})}(t).$

$\mu(t)$  zeigt also den Index  $m$  an, welcher zu dem  $S_m$  gehört, das in  $t$  den Zustand von  $X$  festlegt. Dies entspricht gerade der Anzahl der Sprünge von  $X$  bis  $t$ .  $\tilde{D}$  bezeichnet die bisherige Verweildauer von  $X$  in seinem aktuellen Zustand. Wir wollen zunächst ein eher technisches Lemma über die Häufigkeit von Sprüngen von  $\mu$  beweisen, welches wir wiederholt zur Abschätzung in späteren Beweisen anwenden können.

### Lemma 3.2.2

Es existiert eine von  $k \in \mathbb{N}$ ,  $r \in \mathbb{S}$ ,  $d > 0$  und  $t - d > 0$  unabhängige Konstante  $C$  derart, dass für hinreichend kleines  $h$  gilt

$$P(\mu(t+h) - \mu(t) \geq k | S_m = r, Z_m = t-d, T_{m+1} > d) \leq C^k h^k.$$

**Beweis:** Die Wahrscheinlichkeit lässt sich berechnen durch

$$\begin{aligned} & P(\mu(t+h) - \mu(t) \geq k | S_m = r, Z_m = t-d, T_{m+1} > d) \\ &= \frac{P\left(\sum_{j=1}^k T_{m+j} \leq d+h, T_{m+1} > d \mid S_m = r, Z_m = t-d\right)}{P(T_{m+1} > d | S_m = r, Z_m = t-d)} \\ &= \frac{1}{S_{t-d}^r(d)} \int_d^{d+h} \int_0^{d+h-t_1} \cdots \int_0^{d+h-t_1-\dots-t_{k-1}} \sum_{\substack{s_1, \dots, s_k \in \mathbb{S} \\ s_j \neq s_{j+1}}} f_{t-d}^{rs_1}(t_1) \prod_{j=2}^k f_{t-d+t_1+\dots+t_{j-1}}^{s_{j-1}s_j}(t_j) dt_k \dots dt_1 \\ &\leq \int_d^{d+h} \int_0^{d+h-t_1} \cdots \int_0^{d+h-t_1-\dots-t_{k-1}} \sum_{\substack{s_1, \dots, s_k \in \mathbb{S} \\ s_j \neq s_{j+1}}} \frac{f_{t-d}^{rs_1}(t_1)}{S_{t-d}^r(t_1)} \prod_{j=2}^k f_{t-d+t_1+\dots+t_{j-1}}^{s_{j-1}s_j}(t_j) dt_k \dots dt_1 \\ &\leq |\mathbb{S}|^k \tilde{C}^k h^k =: C^k h^k, \end{aligned}$$

wobei  $\tilde{C}$  hier die Konstante aus (L) bezeichnet. Die erste Abschätzung gilt aufgrund von  $f_{t-d}^{rs_1}(t_1)/S_{t-d}^r(d) \leq f_{t-d}^{rs_1}(t_1)/S_{t-d}^r(t_1)$ , wegen  $t_1 \geq d$  und der Monotonie von  $S$ .  $\square$

Wir beweisen nun ein Lemma, welches wir zugleich nutzen, um Übergangsintensitäten vom Zustand  $r$  in den Zustand  $s$ ,  $\alpha_{rs}$ , zu definieren.

**Lemma 3.2.3**

Unter den bisherigen Definitionen und Annahmen, insbesondere (L), gilt für  $\{r \neq s\} \subset \mathbb{S}$  und  $t > d > 0$

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} P(X(t+h) = s | X(t) = r, \tilde{D}(t) = d, X(u), u \leq t) \\ = \frac{f_{t-d}^{rs}(d)}{S_{t-d}^r(d)} =: \alpha_{rs}(t, d). \end{aligned}$$

**Beweis:** Wir stellen zunächst fest, dass

$$\{\mu(t) = m\} \in \sigma(X(u), u \leq t) \text{ und } \{\mu(t) = m\} \in \sigma(S_0, Z_0, \dots, S_m, Z_m, \mathbb{1}_{\{Z_{m+1} > t\}}). \quad (3.9)$$

Wir können außerdem sehen, dass

$$\sigma((X(u), u \leq t) |_{\{\mu(t)=m\}}) = \sigma((S_0, Z_0, \dots, S_m, Z_m, \mathbb{1}_{\{Z_{m+1} > t\}}) |_{\{\mu(t)=m\}}), \quad (3.10)$$

denn für alle  $\omega \in \{\mu(t) = m\}$  gilt einerseits

$$X(u, \omega) = \sum_{j=0}^m \mathbb{1}_{[Z_j(\omega), Z_{j+1}(\omega))}(u) S_j(\omega), \quad u \leq t,$$

und andererseits

$$\begin{aligned} Z_0(\omega) &= 0, \\ S_0(\omega) &= X(0, \omega), \\ Z_{j+1}(\omega) &= \inf \left\{ q \mathbb{1}_{\{q > Z_j(\omega), X(q, \omega) \neq S_j(\omega)\}} + \infty (\mathbb{1}_{\{q \leq Z_j(\omega)\}} + \mathbb{1}_{\{X(q, \omega) = S_j(\omega)\}}) \mid q \in \mathbb{Q} \right\}, \\ &\quad j = 0, \dots, m-1, \\ S_{j+1}(\omega) &= X(Z_{j+1}(\omega)), \quad j = 0, \dots, m-1, \\ \mathbb{1}_{\{Z_{m+1} > t\}}(\omega) &\equiv 1. \end{aligned}$$

Lemma 3.0.3 liefert dann (3.10). Dies nutzen wir nun für folgende Umformung:

$$\begin{aligned} &P(X(t+h) = s | X(t) = r, \tilde{D}(t) = d, X(u), u \leq t) |_{\{\mu(t)=m\}} \\ &\stackrel{a.}{=} P(X(t+h) = s | S_0, Z_0, \dots, S_{m-1}, Z_{m-1}, S_m = r, Z_m = t-d, T_{m+1} > d) |_{\{\mu(t)=m\}} \\ &\stackrel{b.}{=} \sum_{k=1}^{\infty} P \left( \sum_{j=1}^k T_{m+j} \leq d+h, \sum_{j=1}^{k+1} T_{m+j} > d+h, S_{m+k} = s \mid S_m = r, Z_m = t-d, T_{m+1} > d \right) \end{aligned} \quad (3.11)$$

a. (3.10) und (3.9) erlauben uns die Anwendung von Lemma 3.0.2.

- b. Wir bilden eine abzählbare Zerlegung des Ereignisses  $\{X(t+h) = s, \mu(t) = m, S_m = r\}$  und nutzen die  $\sigma$ -Additivität, um zur Reihe der Wahrscheinlichkeiten überzugehen. Für diese Einzelereignisse macht Definition 3.2.1 klar, dass wir  $S_0, Z_0, \dots, S_{m-1}, Z_{m-1}$  aus der Bedingung entfernen können. Da der neue Term unabhängig von  $\omega$  ist, können wir an dieser Stelle die Mengeneinschränkung unterlassen.

Zunächst untersuchen wir den ersten Summanden dieser Reihe. Die Bedingung  $T_{m+1} > d$  können wir gemäß den Rechenregeln für elementare bedingte Wahrscheinlichkeiten behandeln. Demnach erhalten wir einen Bruch mit dem Nenner

$$P(T_{m+1} > d | S_m = r, Z_m = t - d) = S_{t-d}^r(d) \quad (3.12)$$

und Zähler

$$\begin{aligned} & P(d < T_{m+1} \leq d+h < T_{m+1} + T_{m+2}, S_{m+1} = s | S_m = r, Z_m = t - d) \\ &= \int_d^{d+h} \int_{d+h-t_1}^{\infty} f_{t-d}^{rs}(t_1) f_{t-d+t_1}^s(t_2) dt_2 dt_1 \\ &= \int_d^{d+h} \int_0^{\infty} f_{t-d}^{rs}(t_1) f_{t-d+t_1}^s(t_2) dt_2 dt_1 - \int_d^{d+h} \int_0^{d+h-t_1} f_{t-d}^{rs}(t_1) f_{t-d+t_1}^s(t_2) dt_2 dt_1 \\ &= \int_d^{d+h} f_{t-d}^{rs}(t_1) dt_1 + O(h^2). \end{aligned} \quad (3.13)$$

Die letzte Identität erklärt sich hier einerseits damit, dass sich im ersten Summanden die zweite Dichte zu eins integriert und andererseits im zweiten Summanden der Integrand gemäß (L) beschränkt bleibt. Die Landau- $O$ -Notation bezieht sich hier, wie auch im Folgenden, auf den Grenzübergang  $h \rightarrow 0$ .

Den  $k$ -ten Summanden können wir wegen Lemma 3.2.2 abschätzen durch

$$P(\mu(t+h) - \mu(t) \geq k | S_m = r, Z_m = t - d, T_{m+1} > d) \leq C^k h^k$$

Dies begründet, dass die Reihe ab dem zweiten Summanden in (3.11)  $O(h^2)$  ist, denn

$$\sum_{k=2}^{\infty} C^k h^k = C^2 h^2 \frac{1}{1 - Ch} = O(h^2)$$

für hinreichend kleines  $h$ . Dies – kombiniert mit (3.12) und (3.13), eingesetzt in (3.11) – liefert

$$\begin{aligned} & \frac{1}{h} P(X(t+h) = s | X(t) = r, \tilde{D}(t) = d, X(u), u \leq t) |_{\{\mu(t)=m\}} \\ &= \frac{1}{h} \left\{ \frac{\int_d^{d+h} f_{t-d}^{rs}(t_1) dt_1 + O(h^2)}{S_{t-d}^r(d)} + O(h^2) \right\} \\ &\xrightarrow{h \rightarrow 0} \frac{f_{t-d}^{rs}(d)}{S_{t-d}^r(d)}. \end{aligned}$$

Da dies für alle  $m \in \mathbb{N}$  gilt, und die Mengen  $\{\mu(t) = m\}$  eine Partition von  $\Omega$  bilden, ist das Lemma bewiesen.  $\square$

### 3.2.1 Daten und Schätzer

Wir werden an dieser Stelle nur eine Situation vollständiger Daten diskutieren und Inferenz auf dem Intervall  $[0, 1]$  betreiben. Daher bedeutet „vollständig“ hier, dass  $n$  unabhängig identisch verteilte Beobachtungen von  $X$  auf dem Intervall  $[0, 1]$  vorliegen. Dies ist äquivalent zu  $n$  u.i.v. Ausprägungen der Folge  $((S_m, Z_m) \mathbb{1}_{\{Z_m \leq 1\}})_{m \in \mathbb{N}}$ . Unabhängige Rechtszensierung könnte genau wie im vorigen Kapitel berücksichtigt werden, würde allerdings die Notation weiter verkomplizieren. In einem gewissen Maße ist Zensierung so sogar schon berücksichtigt, indem man einen absorbierenden Zustand „zensiert“ definiert. Die damit verbundenen Annahmen (etwa stetige Verteilung der Zensierungszeit) sind aber zum Teil restriktiver als bei separater Einbeziehung im Stile von Kapitel 3.1.1. Dafür könnte aber von der Annahme der stochastischen Unabhängigkeit von Zensierungszeit und  $X$  etwas abgewichen werden, da z.B. die Übergangsintensität in den Zustand „zensiert“ vom aktuellen Zustand des Prozesses abhängen darf.

Wir wollen nun zu einer Zählprozessnotation übergehen, um den Intensitätsschätzer in der Form von NL definieren und dann theoretisch legitimieren zu können. Um die Notation zu vereinfachen, werden wir zwei Zustände  $\{r \neq s\} \subset \mathbb{S}$  fixieren und diese dann in den Bezeichnungen nicht berücksichtigen. Unser Interesse gilt der Schätzung der Übergangsintensität von  $r$  nach  $s$ , also  $\alpha(t, d) := \alpha_{rs}(t, d)$ ,  $t, d \in [0, 1]$ . Wir definieren

$$\begin{aligned} N(t) &:= \sum_{m=1}^{\infty} \mathbb{1}_{\{Z_m \leq t\}} \mathbb{1}_{\{S_{m-1}=r\}} \mathbb{1}_{\{S_m=s\}}, \\ Y(t) &:= \mathbb{1}_{\{X(t-)=r\}} = \sum_{m=1}^{\infty} \mathbb{1}_{\{Z_{m-1} < t \leq Z_m\}} \mathbb{1}_{\{S_{m-1}=r\}} \quad \text{und} \\ D(t) &:= \tilde{D}(t) Y(t). \end{aligned}$$

Die Interpretationen sind verwandt mit denen des vorangegangenen Kapitels.  $N$  beschreibt die Anzahl der Sprünge von  $r$  nach  $s$  bis zur Zeit  $t$ ,  $Y$  ist ein Indikator, der

anzeigt, ob  $N$  kurz vor  $t$  unter Risiko steht zu springen, und  $D$  zeigt die Dauer an, die  $X$  bereits in Zustand  $r$  verweilt, vorausgesetzt  $X$  ist gerade in  $r$ . Im Unterschied zur vorherigen Situation kann der Zählprozess nun mehr als einmal springen.

Derart vorbereitet können wir den Schätzer  $\hat{\alpha}$  in gleicher Art und Weise wie zuvor definieren:

$$\hat{\alpha}(t, d) := \frac{\sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) dN_i(s)}{\sum_{i=1}^n \int k_b(d - D_i(s)) k_b(t - s) Y_i(s) ds}$$

Die Indizierung erfolgt wieder über die unabhängigen Beobachtungen, welche in der Notation bislang nicht kenntlich gemacht wurden. Wir werden im folgenden Abschnitt zeigen, dass erneut die gesamte Theorie aus Kapitel 2 angewandt werden kann. An dieser Stelle soll aber auf eine wiederholte, auf den Spezialfall bezogene Darstellung verzichtet werden.

Diese Theorie beinhaltet dann insbesondere auch Tests auf die Abhängigkeit der Funktion  $\alpha(t, d)$  vom Argument  $d$ , wie bereits auf Seite 47 für das einfachere Dreizustandsmodell beschrieben wurde. Falls jede der Übergangsintensitäten  $\alpha_{rs}(t, d)$  von  $d$  unabhängig ist, ist das Modell nicht nur semi-markovsch, sondern sogar markovsch.

### 3.2.2 Nachweis der Martingaleigenschaft

Erneut ist zu beweisen, dass die beschriebenen Prozesse in den Modellrahmen passen, unter dem NL und unser Kapitel 2 operieren. Dementsprechend arbeiten wir mit der Filtration  $\mathfrak{F}_t := (N(u), D(u+), Y(u+), u \leq t)$ . Den rechtsseitigen Grenzwert von  $Y$  bzw.  $D$  einzubeziehen, sichert die Rechtsstetigkeit der Filtration und somit die Erfüllung der „conditions habituelles“. Die notwendigen Aussagen fassen wir erneut in einem Satz zusammen.

#### Satz 3.2.4

*Mit den Definitionen und Annahmen aus Kapitel 3.2 ist die Filtration  $\mathfrak{F}_t$  rechtsstetig. Außerdem ist  $\alpha(t, D(t)) \cdot Y(t)$  vorhersagbar und*

$$M(t) := N(t) - \int_0^t \alpha(s, D(s)) Y(s) ds$$

*ein Martingal, jeweils bezüglich  $\mathfrak{F}_t$ . Dies bedeutet insbesondere, dass  $\alpha(t, D(t)) \cdot Y(t)$  der Intensitätsprozess von  $N(t)$  ist.*

Der Beweis des Satzes wird einfach, sofern wir folgendes Lemma beweisen können.

**Lemma 3.2.5**

Es existiert eine von  $t$  und  $\omega$  unabhängige Konstante  $C > 0$  derart, dass

$$|\mathbb{E}[M(t+h) - M(t)|\mathfrak{F}_t]| \leq Ch^2.$$

Wir werden nun zunächst zeigen, wie wir daraus Satz 3.2.4 herleiten können, und uns dann dem Nachweis des Lemmas widmen.

**Beweis von Satz 3.2.4:** Die Vollständigkeit der Filtration und die Vorhersagbarkeit von  $\alpha Y$  folgen analog zum Beweis von Satz 3.1.3.

Lemma 3.2.5 ermöglicht den Nachweis der Martingaleigenschaft von  $M$ :

$$\begin{aligned} |\mathbb{E}[M(t+h) - M(t)|\mathfrak{F}_t]| &\leq \sum_{j=1}^k \left| \mathbb{E} \left[ M \left( t + \frac{jh}{k} \right) - M \left( t + \frac{(j-1)h}{k} \right) \middle| \mathfrak{F}_t \right] \right| \\ &= \sum_{j=1}^k \left| \mathbb{E} \left[ \mathbb{E} \left[ M \left( t + \frac{jh}{k} \right) - M \left( t + \frac{(j-1)h}{k} \right) \middle| \mathfrak{F}_{t+\frac{(j-1)h}{k}} \right] \middle| \mathfrak{F}_t \right] \right| \\ &\leq \sum_{j=1}^k C \left( \frac{1}{k} \right)^2 = \frac{C}{k}. \end{aligned} \quad (3.14)$$

Der Ausdruck  $|\mathbb{E}[M(t+h) - M(t)|\mathfrak{F}_t]|$  – selbst unabhängig von  $k$  – ist also durch  $C/k$  beschränkt für beliebige  $k$  und muss daher identisch null sein. Damit ist auch  $\mathbb{E}[M(t+h) - M(t)|\mathfrak{F}_t]$  selbst identisch null und die Martingaleigenschaft gezeigt – vorausgesetzt wir können Lemma 3.2.5 beweisen.  $\square$

Für den Nachweis des Lemmas 3.2.5 definieren wir das Zufallselement  $\mathcal{F}_t := (N(u), D(u+), Y(u+), u \leq t)$ , sodass also gilt  $\mathfrak{F}_t = \sigma(\mathcal{F}_t)$ .  $\mathcal{F}_t$  ist damit ein Zufallselement mit Werten im Raum der Funktionen  $\{f : [0, t] \rightarrow \mathbb{R}^3\}$ . Die adäquate  $\sigma$ -Algebra auf diesem Raum ist die von den Zylindermengen erzeugte. Wir wollen diese aber nicht in der Notation explizit machen. Da offensichtlich  $D(t)$  und  $Y(t)$   $\mathcal{F}_t$ -messbar sind, existieren gemäß dem Faktorisierungslemma messbare Funktionen  $d_t$  und  $y_t$ , sodass  $D(t) = d_t(\mathcal{F}_t)$  und  $Y(t) = y_t(\mathcal{F}_t)$ . Das folgende Lemma wird eine tragende Rolle im Beweis von Lemma 3.2.5 spielen.

**Lemma 3.2.6**

Mit den genannten Definitionen erhalten wir

(a)

$$\begin{aligned} & \mathbb{E} [M(t+h) - M(t) | \mathcal{F}_t = f] \Big|_{\{f|y_t(f)=1\}} \\ &= \mathbb{E} [M(t+h) - M(t) | Z_m = t - d_t(f), T_{m+1} > d_t(f), S_m = r] \end{aligned}$$

für  $P_{\mathcal{F}_t}$ -f.a.  $f$  und beliebiges  $m \in \mathbb{N}$  sowie

(b)

$$\begin{aligned} & \left| \mathbb{E} [M(t+h) - M(t) | \mathcal{F}_t = f] \right| \Big|_{\{f|y_t(f)=0\}} \\ & \leq \sup_{\tilde{r} \in \mathbb{S} \setminus \{r\}, z \in [0,1]} \mathbb{E} [N(t+h) - N(t) | Z_m = z, T_{m+1} > t - z, S_m = \tilde{r}] \\ & + \sup_{\tilde{r} \in \mathbb{S} \setminus \{r\}, z \in [0,1]} \mathbb{E} \left[ \int_t^{t+h} \alpha(u, D(u)) Y(u) du \Big| Z_m = z, T_{m+1} > t - z, S_m = \tilde{r} \right] \end{aligned}$$

für ebenfalls  $P_{\mathcal{F}_t}$ -f.a.  $f$  und beliebiges  $m \in \mathbb{N}$ .

**Beweis:** (a) Im Folgenden bezeichnen wir  $\Delta M := M(t+h) - M(t)$ ,  $\Delta N := N(t+h) - N(t)$  und  $\Delta A := \int_t^{t+h} \alpha(u, D(u)) Y(u) du$ . Wir beginnen mit dem Nachweis folgender Behauptung:

*Behauptung 1:* Für jedes  $t \in [0, 1]$  und jedes  $r \in \mathbb{S}$  existiert eine von  $m$  sowie  $s_0, z_0, \dots, s_{m-1}, z_{m-1}$  unabhängige Funktion  $^M g_t^r(z)$  mit

$$\mathbb{E} [\Delta M | S_0 = s_0, Z_0 = z_0, \dots, S_m = r, Z_m = z, T_{m+1} > t - z] = ^M g_t^r(z).$$

Analog existieren auch Funktionen  $^N g_t^r(z)$  und  $^A g_t^r(z)$  für  $\Delta N$  und  $\Delta A$  anstelle von  $\Delta M$ .

*Beweis von Behauptung 1:* Wegen  $\Delta M = \Delta N - \Delta A$  reicht der Nachweis für jeweils  $\Delta N$  und  $\Delta A$ . Wir betrachten zunächst  $\Delta N$ . Gemäß der Definition von  $N$  ist leicht zu erkennen, dass  $\Delta N = \sum_{\nu=0}^{\infty} \mathbb{1}_{\{S_\nu=r\}} \mathbb{1}_{\{S_{\nu+1}=s\}} \mathbb{1}_{\{t < Z_{\nu+1} \leq t+h\}}$ . Für die einzelnen Summanden erhalten wir für  $\nu \geq m$

$$\begin{aligned} & \mathbb{E} [\mathbb{1}_{\{S_\nu=r\}} \mathbb{1}_{\{S_{\nu+1}=s\}} \mathbb{1}_{\{t < Z_{\nu+1} \leq t+h\}} | S_0 = s_0, Z_0 = z_0, \dots, S_m = r, Z_m = z, T_{m+1} > t - z] \\ &= P(S_\nu = r, S_{\nu+1} = s, Z_{\nu+1} \in (t, t+h] | S_m = r, Z_m = z, T_{m+1} > t - z) \\ &= \int_{t-z}^{t-z+h} \int_{t-z+t_1}^{t-z+t_1+h} \dots \int_{t-z+t_1+\dots+t_{\nu-m}}^{t-z+t_1+\dots+t_{\nu-m}+h} \sum_{\substack{s_1, \dots, s_{\nu-m-1} \in \mathbb{S} \\ s_j \neq s_{j+1}}} f_z^{rs_1}(t_1) \prod_{j=2}^{\nu-m-1} f_{z+t_1+\dots+t_{j-1}}^{s_{j-1}s_j}(t_j) \cdot \\ & \quad f_{z+t_1+\dots+t_{\nu-m-1}}^{s_{\nu-m-1}r}(t_{\nu-m}) f_{z+t_1+\dots+t_{\nu-m}}^{rs}(t_{\nu-m+1}) dt_{\nu-m+1} \dots dt_1. \end{aligned} \quad (3.15)$$



Die erste Identität gilt gemäß Definition 3.2.1. Der letzte Term ist nur von  $\nu - m$  abhängig, nicht aber von  $m$  selbst. Für  $\nu < m$  lautet der bedingte Erwartungswert null. Wenn wir also die rechte Seite von (3.15) mit  ${}^{\nu-m}h_t^r(z)$  bezeichnen, ergibt sich

$$\mathbb{E}[N(t+h) - N(t) | S_0 = s_0, Z_0 = z_0, \dots, S_m = r, Z_m = z, T_{m+1} > t - z] = \sum_{j=0}^{\infty} {}^jh_t^r(z),$$

wobei die Reihe auf der rechten Seite die Eigenschaften der für Behauptung 1 gesuchten Funktion  ${}^Ng_t^r(z)$  erfüllt. Eine ähnliche Argumentation liefert ein analoges Ergebnis für  $\Delta A$ . Hier kann die Darstellung  $\Delta A = \sum_{\nu=0}^{\infty} \int_{t \vee Z_\nu}^{(t+h) \wedge Z_\nu} \mathbb{1}_{\{S_\nu=r\}} \alpha(u, u - Z_\nu) d\nu$  verwendet werden. Dies vollendet den Beweis von *Behauptung 1*.

Diese Aussage können wir in folgender Gleichungskette anwenden: Es sei  $B \in \sigma(\mathcal{F}_t) \cap \mathfrak{P}(\{Y(t) = 1\})$  derart, dass eine messbare Teilmenge  $\mathcal{F}_t(B)$  des Bildraumes von  $\mathcal{F}_t$  existiert, deren Urbild  $B$  ist.

$$\begin{aligned} \int_{\mathcal{F}_t(B)} \mathbb{E}[\Delta M | \mathcal{F}_t = f] dP_{\mathcal{F}_t}(f) &\stackrel{a.}{=} \int_B \mathbb{E}[\Delta M | \mathcal{F}_t] dP \\ &\stackrel{b.}{=} \int_B \Delta M dP \stackrel{c.}{=} \sum_{m=0}^{\infty} \int_{B \cap \{\mu(t)=m\}} \Delta M dP \\ &\stackrel{d.}{=} \sum_{m=0}^{\infty} \int_{B \cap \{\mu(t)=m\}} \mathbb{E}[\Delta M | S_0, Z_0, \dots, S_m, Z_m, \mathbb{1}_{\{Z_{m+1}>t\}}] dP \\ &\stackrel{e.}{=} \sum_{m=0}^{\infty} \int_{B \cap \{\mu(t)=m\}} {}^Mg_t^r(Z_m) dP \\ &\stackrel{f.}{=} \sum_{m=0}^{\infty} \int_{B \cap \{\mu(t)=m\}} {}^Mg_t^r(t - D(t)) dP \\ &\stackrel{g.}{=} \int_B {}^Mg_t^r(t - D(t)) dP \\ &\stackrel{h.}{=} \int_{\mathcal{F}_t(B)} {}^Mg_t^r(t - d_t(f)) dP_{\mathcal{F}_t}(f) \end{aligned}$$

- a. Maß-Transformationssatz.
- b. Definition des bedingten Erwartungswertes, wegen  $B \in \sigma(\mathcal{F}_t)$ .
- c. Die Ereignisse  $\{\mu(t) = m\}$  bilden eine Zerlegung von  $\Omega$ .
- d. Folgt nach der Definition bedingter Erwartungswerte, denn  $B \cap \{\mu(t) = m\}$  ist ein  $(S_0, Z_0, \dots, S_m, Z_m, \mathbb{1}_{\{Z_{m+1}>t\}})$ -messbares Ereignis. Dies ist daran zu erkennen, dass für alle  $\omega \in \{\mu(t) = m\}$  die kompletten Pfade von  $\mathcal{F}_t$  rekonstruierbar sind.
- e. Für alle  $\omega \in B \cap \{\mu(t) = m\}$  gelten  $S_m = r$  und  $\mathbb{1}_{\{Z_{m+1}>t\}} = 1 \Leftrightarrow T_{m+1} > t - Z_m$ . Da der bedingte Erwartungswert die Komposition von faktorisiertem

bedingtem Erwartungswert und der bedingenden Zufallsvariable darstellt, können wir Behauptung 1 anwenden, um die Funktion  ${}^M g_t^r$  einzusetzen.

- f. Für alle  $\omega \in B \cap \{\mu(t) = m\}$  ist  $Z_m = t - D(t)$ .
- g. Da der Integrand unabhängig von  $m$  ist, können wir wieder zur Integration über gesamt  $B$  zurückkehren.
- h. Maß-Transformationssatz.

Da diese Rechnung für alle  $B \in \sigma(\mathcal{F}_t) \cap \mathfrak{P}(\{Y(t) = 1\})$  und damit für alle entsprechenden Bildmengen  $\mathcal{F}_t(B)$  gilt, ist Teil (a) gezeigt (indem man etwa Elstrodt, 2009, Kapitel IV, Satz 4.4, anwendet).

(b) Zunächst gilt  $|\Delta M| \leq \Delta N + \Delta A$ . Wir stellen außerdem fest, dass die rechte Seite der Aussage von (b)  $\sup_{\tilde{r} \neq r, z} {}^N g_t^{\tilde{r}}(z) + \sup_{\tilde{r} \neq r, z} {}^A g_t^{\tilde{r}}(z)$  entspricht. Wir argumentieren zunächst ähnlich wie im Beweis von Teil (a) und wählen diesmal  $B \in \sigma(\mathcal{F}_t) \cap \mathfrak{P}(\{Y(t) = 0\})$  derart, dass eine messbare Teilmenge  $\mathcal{F}_t(B)$  des Bildraumes von  $\mathcal{F}_t$  existiert, deren Urbild  $B$  ist.

$$\begin{aligned}
\int_{\mathcal{F}_t(B)} \mathbb{E}[\Delta N | \mathcal{F}_t = f] dP_{\mathcal{F}_t}(f) &= \int_B \mathbb{E}[\Delta N | \mathcal{F}_t] dP \\
&= \int_B \Delta N dP = \sum_{m=0}^{\infty} \int_{B \cap \{\mu(t)=m\}} \Delta N dP \\
&= \sum_{m=0}^{\infty} \sum_{\tilde{r} \neq r} \int_{B \cap \{\mu(t)=m\} \cap \{S_m=\tilde{r}\}} \mathbb{E}[\Delta N | S_0, Z_0, \dots, S_m, Z_m, \mathbb{1}_{\{Z_{m+1}>t\}}] dP \\
&= \sum_{m=0}^{\infty} \sum_{\tilde{r} \neq r} \int_{B \cap \{\mu(t)=m\} \cap \{S_m=\tilde{r}\}} {}^N g_t^{\tilde{r}}(Z_m) dP \\
&\leq \sum_{m=0}^{\infty} \sum_{\tilde{r} \neq r} \int_{B \cap \{\mu(t)=m\} \cap \{S_m=\tilde{r}\}} \sup_{\tilde{r} \neq r, z} {}^N g_t^{\tilde{r}}(z) dP \\
&= \int_B \sup_{\tilde{r} \neq r, z} {}^N g_t^{\tilde{r}}(z) dP = \int_{\mathcal{F}_t(B)} \sup_{\tilde{r} \neq r, z} {}^N g_t^{\tilde{r}}(z) dP_{\mathcal{F}_t}(f)
\end{aligned}$$

Die Argumente der ersten vier Zeilen sind identisch zu denen in Teil (a). Dann können wir den Integranden durch das deterministische und von  $m$  unabhängige Supremum abschätzen. Die letzten Schritte beinhalten wieder das Zusammenfassen der Mengen, über die integriert wird, und die Anwendung des Transformationssatzes. Insgesamt erhalten wir dadurch  $\mathbb{E}[\Delta N | \mathcal{F}_t = f] \leq \sup_{\tilde{r} \neq r, z} {}^N g_t^{\tilde{r}}(z), P_{\mathcal{F}_t}(f)$ -f.s. Auf gleichem Wege erhalten wir ein analoges Resultat für  $\Delta A$  und der Beweis ist vollendet.  $\square$

Derart ausgerüstet widmen wir uns dem Nachweis von Lemma 3.2.5.

**Beweis von Lemma 3.2.5:** Wir zeigen vier Teilaussagen. Alle gelten jeweils für hinreichend kleine  $h$ .

*Aussage 1:* Es existiert eine von  $t$  und  $d$  unabhängige Konstante  $C$  derart, dass

$$\mathbb{E}[\Delta N | Z_m = t - d, S_m = r, T_{m+1} > d] = \frac{\int_d^{d+h} f_{t-d}^{rs}(u) du}{S_{t-d}^r(d)} + R_N(t, d, h),$$

mit  $|R_N(t, d, h)| \leq Ch^2$ .

*Aussage 2:* Es existiert eine von  $t$  und  $d$  unabhängige Konstante  $C$  derart, dass

$$\mathbb{E}[\Delta A | Z_m = t - d, S_m = r, T_{m+1} > d] = \frac{\int_d^{d+h} f_{t-d}^{rs}(u) du}{S_{t-d}^r(d)} + R_A(t, d, h),$$

mit  $|R_A(t, d, h)| \leq Ch^2$ .

*Aussage 3:* Es existiert eine von  $t$  und  $d$  unabhängige Konstante  $C$  derart, dass für alle  $\tilde{r} \in \mathbb{S} \setminus \{r\}$

$$\mathbb{E}[\Delta N | Z_m = t - d, S_m = \tilde{r}, T_{m+1} > d] \leq Ch^2.$$

*Aussage 4:* Es existiert eine von  $t$  und  $d$  unabhängige Konstante  $C$  derart, dass für alle  $\tilde{r} \in \mathbb{S} \setminus \{r\}$

$$\mathbb{E}[\Delta A | Z_m = t - d, S_m = \tilde{r}, T_{m+1} > d] \leq Ch^2.$$

Aussagen 1 und 2 – gemeinsam mit Teil (a) des Lemmas 3.2.6 – beweisen das Lemma auf dem Ereignis  $\{Y(t) = 1\}$ . Aussagen 3 und 4 – gemeinsam mit Teil (b) des Lemmas 3.2.6 – beweisen es auf dem Ereignis  $\{Y(t) = 0\}$ .

*Beweis von Aussage 1:* Zunächst gilt

$$\mathbb{E}[\Delta N | Z_m = t - d, S_m = r, T_{m+1} > d] = \sum_{k=1}^{\infty} P(\Delta N \geq k | Z_m = t - d, S_m = r, T_{m+1} > d). \quad (3.16)$$

Wegen  $\{\Delta N \geq k\} \subset \{\mu(t+h) - \mu(t) \geq k\}$  können wir den  $k$ -ten Summanden gemäß Lemma 3.2.2 durch  $\tilde{C}^k h^k$  abschätzen, wobei  $\tilde{C}$  unabhängig von  $t, d$  ist. Damit ist die Reihe ab dem zweiten Summanden beschränkt durch

$$\sum_{k=2}^{\infty} \tilde{C}^k h^k = \tilde{C}^2 / (1 - \tilde{C}h) h^2 \leq \bar{C} h^2$$

für kleines  $h$ . Für den ersten Summanden in (3.16) erhalten wir

$$\begin{aligned}
& P(\Delta N \geq 1 | Z_m = t - d, S_m = r, T_{m+1} > d) \\
&= P(Z_{m+1} \leq t + h, S_{m+1} = s | Z_m = t - d, S_m = r, T_{m+1} > d) \\
&+ P(„X springt  $\geq 2$ -mal in  $(t, t + h]$ ,  $\geq 1$ -mal von  $r$  nach  $s$ “ |  $Z_m = t - d,$  \\
& $S_m = r, T_{m+1} > d$ ).
\end{aligned} \tag{3.17}$$

Das Ereignis der Wahrscheinlichkeit im zweiten Summanden, welches hier der Einfachheit halber nur in Worten beschrieben wurde, ist eine Teilmenge von  $\{\mu(t + h) - \mu(t) \geq 2\}$  und damit gemäß Lemma 3.2.2 für kleine  $h$  durch  $\tilde{C}h^2$  beschränkt. Es bleibt der erste Summand von (3.17) zu untersuchen.

$$\begin{aligned}
& P(Z_{m+1} \leq t + h, S_{m+1} = s | Z_m = t - d, S_m = r, T_{m+1} > d) \\
&= \frac{P(d < T_{m+1} \leq d + h, S_{m+1} = s | Z_m = t - d, S_m = r)}{P(T_{m+1} > d | Z_m = t - d, S_m = r)} \\
&= \frac{\int_d^{d+h} f_{t-d}^{rs}(u) du}{S_{t-d}^r(d)},
\end{aligned}$$

gemäß den Definitionen von  $f_{t-d}^{rs}$  und  $S_{t-d}^r$ . Dies beweist *Aussage 1*.

*Beweis von Aussage 2:* Wegen

$$\Delta A = \int_t^{t+h} \alpha(u, D(u)) Y(u) du = \sum_{l=0}^{\infty} \int_t^{t+h} \alpha(u, u - Z_l) \mathbf{1}_{(Z_l, Z_{l+1}]}(u) \mathbf{1}_{\{S_l=r\}} du$$

erhalten wir

$$\begin{aligned}
& \mathbb{E}[\Delta A | Z_m = t - d, S_m = r, T_{m+1} > d] \\
&= \sum_{l=0}^{\infty} \int_t^{t+h} \mathbb{E}[\alpha(u, u - Z_l) \mathbf{1}_{(Z_l, Z_{l+1}]}(u) \mathbf{1}_{\{S_l=r\}} | Z_m = t - d, S_m = r, T_{m+1} > d] du \\
&= \sum_{l=m}^{\infty} \int_t^{t+h} \mathbb{E}[\alpha(u, u - Z_l) \mathbf{1}_{(Z_l, Z_{l+1}]}(u) \mathbf{1}_{\{S_l=r\}} | Z_m = t - d, S_m = r, T_{m+1} > d] du,
\end{aligned} \tag{3.18}$$

da eine der Indikatorfunktionen gegeben die Bedingung null ist für  $l < m$ . Für  $l > m$  kann der  $l$ -te Summand gemäß (L) unter Ausnutzung von Lemma 3.2.2 abgeschätzt

werden durch

$$\begin{aligned}
& \int_t^{t+h} \mathbb{E}[\alpha(u, u - Z_l) \mathbb{1}_{(Z_l, Z_{l+1}]}(u) \mathbb{1}_{\{S_l=r\}} | Z_m = t - d, S_m = r, T_{m+1} > d] du \\
& \leq C_1 \int_t^{t+h} P(Z_l \leq u | Z_m = t - d, S_m = r, T_{m+1} > d) du \\
& \leq C_1 \int_t^{t+h} P(\mu(u) - \mu(t) > l - m | Z_m = t - d, S_m = r, T_{m+1} > d) du \\
& \leq C_1 \int_t^{t+h} \bar{C}_2^{l-m} h^{l-m} du \leq \tilde{C}^{l-m+1} h^{l-m+1}
\end{aligned} \tag{3.19}$$

für kleine  $h$ . Demnach kann die Reihe der Summanden von (3.18) ab  $l = m + 1$  für kleine  $h$  abgeschätzt werden durch

$$\sum_{l=m+1}^{\infty} \tilde{C}^{l-m+1} h^{l-m+1} \leq \frac{\tilde{C}^2}{1 - \tilde{C}h} h^2 \leq \bar{C} h^2.$$

Für den ersten Summanden, d.h. für  $l = m$ , erhalten wir

$$\begin{aligned}
& \int_t^{t+h} \mathbb{E}[\alpha(u, u - Z_m) \mathbb{1}_{(Z_m, Z_{m+1}]}(u) \mathbb{1}_{\{S_m=r\}} | Z_m = t - d, S_m = r, T_{m+1} > d] du \\
& \stackrel{a.}{=} \int_t^{t+h} \alpha(u, u - t + d) P(T_{m+1} > d - t + u | Z_m = t - d, S_m = r, T_{m+1} > d) du \\
& \stackrel{b.}{=} \int_t^{t+h} \frac{\int_t^{t+h} f_{t-d}^{rs}(u - t + d)}{S_{t-d}^r(u - t + d)} \frac{S_{t-d}^r(u - t + d)}{S_{t-d}^r(d)} du \\
& \stackrel{c.}{=} \frac{\int_d^{d+h} f_{t-d}^{rs}(u) du}{S_{t-d}^r(d)}.
\end{aligned}$$

- a. Es gilt  $\mathbb{1}_{(Z_m, Z_{m+1}]}(u) = \mathbb{1}_{\{Z_m \leq u\}} \mathbb{1}_{\{T_{m+1} > t - Z_m\}}$ . Alle bezüglich der Bedingung messbaren Faktoren werden aus dem bedingten Erwartungswert gezogen, wobei die beiden herausziehbaren Indikatorfunktionen identisch eins sind. Der im bedingten Erwartungswert verbleibende Indikator wird zu einer bedingten Wahrscheinlichkeit umgeschrieben.
- b. Die Definition von  $\alpha$ , siehe Lemma 3.2.3, wird eingesetzt. Außerdem gilt

$$\begin{aligned}
& P(T_{m+1} > d - t + u | Z_m = t - d, S_m = r, T_{m+1} > d) \\
& = \frac{P(T_{m+1} > d - t + u | Z_m = t - d, S_m = r)}{P(T_{m+1} > d | Z_m = t - d, S_m = r)},
\end{aligned}$$

was mit der Definition von  $S_{t-d}^r$  die Umformung erklärt.

- c. Kürzen sowie lineare Substitution.

Damit ist auch Aussage 2 gezeigt.

*Beweis von Aussage 3:* Wir beobachten, dass für  $\tilde{r} \neq r$  wegen Lemma 3.2.2

$$\begin{aligned} P(\Delta N \geq k | Z_m = t - d, S_m = \tilde{r}, T_{m+1} > d) \\ \leq P(\mu(t + h) - \mu(t) \geq k + 1 | Z_m = t - d, S_m = \tilde{r}, T_{m+1} > d) \leq \tilde{C}^{k+1} h^{k+1}, k \geq 1. \end{aligned}$$

Dies begründet sich dadurch, dass, falls  $S_m \neq r$ , der Prozess  $X$  – und damit  $\mu$  – mindestens  $2k \geq k + 1$  springen muss, damit unter diesen  $k$  Sprünge vom Zustand  $r$  in den Zustand  $s$  – also Zuwächse von  $N$  – sein können. Wegen (3.16) mit  $\tilde{r}$  anstatt  $r$  ergibt sich somit für kleine  $h$

$$\mathbb{E}[\Delta N | Z_m = t - d, S_m = \tilde{r}, T_{m+1} > d] \leq \sum_{k=1}^{\infty} \tilde{C}^{k+1} h^{k+1} = \frac{\tilde{C}^2 h^2}{1 - \tilde{C}h} \leq Ch^2.$$

Dies zeigt *Aussage 3*.

*Beweis von Aussage 4:* Wir können die Argumentation aus dem Beweis von Aussage 2 wiederholen mit dem Unterschied, dass der Summand  $l = m$  in (3.18) dieses Mal identisch null ist, da  $\mathbb{1}_{\{S_m=r\}}$  gegeben die Bedingungen konstant null ist. Dies genügt als Beweis von *Aussage 4*, womit auch der Beweis von Lemma 3.2.5 abgeschlossen ist.  $\square$

# Kapitel 4

## Anwendung auf Daten

### 4.1 Die Daten

Wir wollen die vorgestellten Methoden anhand eines Datensatzes vorführen. Der R-Code, welcher allen Berechnungen zu Grunde liegt, befindet sich im Appendix. Wir untersuchen Daten der deutschen Krankenkasse AOK, welche im Zeitraum zwischen 2004 und 2015 gesammelt wurden (siehe auch Weißbach u. a., 2021). Für insgesamt 230.922 Personen liegen Daten vor. Diese umfassen jeweils den Geburtszeitpunkt, die Demenz-Diagnosen während des Beobachtungszeitraumes sowie den Todeszeitpunkt, falls dieser während der Beobachtung eintrat. Die jüngste Person war zu Beginn der Beobachtung 2004 50 Jahre alt, die älteste 106. Die Angaben zu Geburts- und Todeszeitpunkt liegen monatsgenau vor, für Demenz-Diagnosen ist jeweils nur das Quartal bekannt. Knapp 14% der beobachteten Personen erhalten während der Zeit eine Demenzdiagnose, dies entspricht knapp 32.000 Versicherten. Einige der Personen verlassen während des Beobachtungszeitraumes die Beobachtungsmenge, in diesen Fällen liegt Zensierung vor (dies betrifft gut 15% der Personen). Ebenfalls liegt Zensierung bei allen Personen vor, die am Ende des Beobachtungszeitraumes noch am Leben sind, was etwas mehr als 52% der Versicherten sind. Frauen sind im Datensatz etwas häufiger vertreten als Männer (56%) und auch relativ häufiger von Demenz betroffen. Siehe Abbildung 4.1. Da sich die beiden Geschlechter merklich in ihrer Lebenserwartung unterscheiden (Frauen sterben im Schnitt später und erhalten auch später Demenzdiagnosen), erscheint es sinnvoll, unsere Analysen getrennt nach Geschlechtern durchzuführen. Um Wiederholungen zu vermeiden werden hier nur Ergebnisse der Analyse von Daten weiblicher Personen vorgestellt. Die Resultate basieren dementsprechend auf Daten von insgesamt 130.168 Frauen, von denen 20.721 im Studienverlauf eine Demenzdiagnose erhielten.

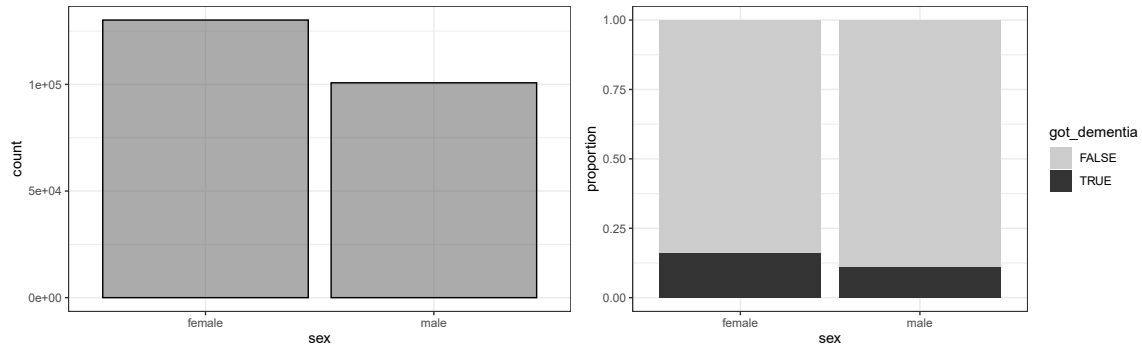


Abbildung 4.1: Geschlechtsverteilung im Datensatz und jeweiliger Anteil an Demenzerkrankungen.

## 4.2 Das Modell

Wir modellieren diese Daten in Form eines Drei-Zustands-Modells – „gesund“, „dement“ und „tot“. Vom Zustand „gesund“ kann in beide anderen Zustände übergegangen werden, von „dement“ nur in „tot“, der Zustand „tot“ ist absorbierend. Es handelt sich also um ein in der Epidemiologie übliches unidirektionales Krankheit-Tod-Modell (siehe etwa Andersen u. Keiding, 2002, Kapitel 3.3). Abbildung 4.2 illustriert die Struktur. Dieses Modell ist ein Semi-Markov-Modell im Sinne von Kapitel 3.2, weshalb die Theorie des Kapitels anwendbar ist.

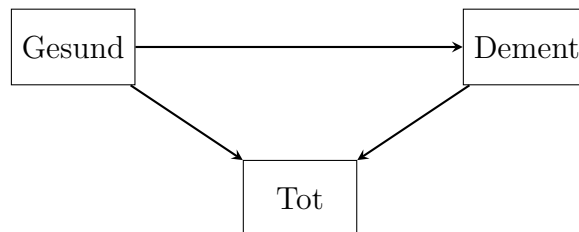


Abbildung 4.2: Strukturdiagramm des Dreizustandsmodells.

Unser Interesse gilt der Übergangsintensität vom Zustand „dement“ in den Zustand „tot“, welche wir mit  $\alpha(t, d)$  bezeichnen. Die Zeitachse  $t$  bezeichnet hier das Alter, während die Achse  $d$  die Dauer seit der Demenzerkrankung misst.

Für die Schätzung der Übergangsintensität relevant sind nur Daten von Personen, welche zu irgendeinem Zeitpunkt eine Demenzerkrankung erhalten haben. Für diese können wir die Daten auf das in Kapitel 3.1.1 beschriebene Format der Daten  $(Z_1, Z_2, \Delta)$  transformieren, um damit die Schätzer berechnen und Tests durchführen zu können. Hier entspricht  $Z_1$  dem Alter bei Demenzerkrankung und  $Z_2$  dem Minimum aus Alter bei Tod und Alter bei Zensurierung.  $\Delta$  zeigt an, ob Zensurierung vorliegt. Für alle Untersuchungen wurde ein Dreieckskern  $k(x) = \mathbb{1}_{[-1,1]}(x)(1 - |x|)$  genutzt.

**Bemerkung:** Obwohl Datenpunkte von Personen, die nie eine Demenzerkrankung erhalten, den Schätzer nicht beeinflussen, ist – bei präziser Zuordnung der Notation von Kapitel



3.2 auf unsere Daten – der Stichprobenumfang  $n$  gleich der Anzahl der Frauen im Datensatz (130.168) und nicht gleich der Frauen mit Demenzdiagnose (20.721). Der Schätzer selbst wird dadurch nicht beeinflusst, denn für diese Personen ist sowohl  $dN_i \equiv 0$  als auch  $Y_i \equiv 0$ . Der Varianzschätzer  $\hat{\sigma}_x^2$  gemäß Satz 2.1.4 verändert sich allerdings. Dies macht insofern Sinn, als dass die asymptotische Varianz  $\sigma_x^2$  den Faktor  $1/P(Y(t) = 1)$  enthält. Die Stichprobe auf Personen zu beschränken, die irgendwann eine Demenzdiagnose erhalten, entspricht einer bedingten Betrachtung gegeben eben diese Bedingung. Und die Wahrscheinlichkeit, dass eine Person im Alter  $t$  Demenz hat (also  $P(Y(t) = 1)$ ) ist gegeben, dass sie irgendwann eine Demenzdiagnose erhält, offensichtlich größer. Der Standardfehler des Schätzers wird allerdings nicht beeinflusst. Wenn wir den Varianzschätzer mit dem Faktor  $1/(nb^2)$  renormieren, um den Standardfehler zu erhalten, kürzt sich der Faktor  $n$ , welcher sich zuvor verändert hatte, heraus, und wir erhalten das selbe Ergebnis wie mit der alternativen Fallzahl (weil auch im Varianzschätzer die Summen im Zähler und Nenner von den Datenpunkten von Personen, die nie eine Diagnose erhalten, unberührt bleiben).

### 4.3 Ergebnisse der Intensitätsschätzung

Das Thema der Bandbreitenwahl war nicht Teil dieser Arbeit und soll auch hier nicht ausführlich diskutiert werden. Wir stellen hier Ergebnisse der Schätzung für eine feste, nach eigenem Ermessen gewählte Kombination von Bandbreiten vor. Für die  $t$ -Achse lautet diese 2 Jahre. Da der relevante Abschnitt der Skala auf der  $d$ -Achse kürzer ist, wählen wir die Bandbreite in dieser Richtung ebenfalls kleiner, nämlich als 1,33 Jahre. Abbildung 4.3 stellt die resultierende Schätzung der Intensitätsfunktion  $\alpha(t, d)$  dar. Es sei darauf hingewiesen, dass alle Abbildungen Schätzungen monatlicher Hazardraten anzeigen, obwohl die Achsen in Jahren skaliert sind. Die Form der Funktionen ist davon nicht betroffen, einzig das Niveau der Werte muss für eine Umrechnung gegebenenfalls mit dem Faktor zwölf reskaliert werden. Für jedes feste  $d$  ist ein etwa exponentieller Anstieg in  $t$ -Richtung zu erkennen. In  $d$ -Richtung ist die Schätzung zunächst hoch, fällt dann recht steil ab, um anschließend wieder anzusteigen. Für große  $d$  ist der Trend je nach  $t$  unterschiedlich, in diesem Bereich liegen auch weniger Daten vor, sodass die Schätzung als weniger zuverlässig einzuschätzen ist. Dieses Muster kann auch in Abbildung 4.4 erkannt werden, welche drei Querschnitte der geschätzten Funktion für die festen Werte  $t = 70, 80, 90$  Jahre zeigt. Ebenfalls sind zum Vergleich Schätzungen der Hazardrate unter der Markov-Annahme eingezeichnet, also der Annahme, dass die Funktion  $\alpha(t, d)$  von  $d$  unabhängig ist. Im nächsten Kapitel soll diese Annahme statistisch getestet werden.

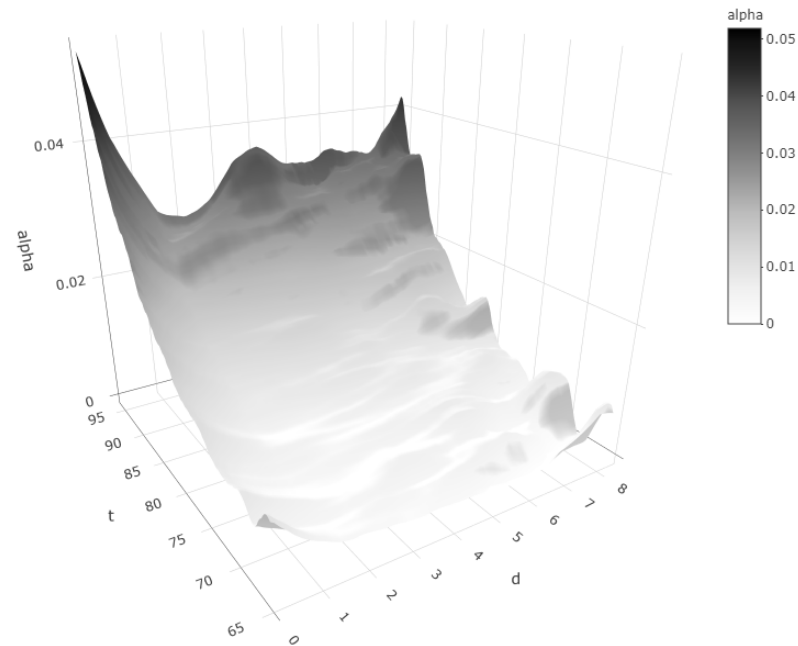


Abbildung 4.3: Geschätzte Übergangsintensität  $\hat{\alpha}(t, d)$  für Bandbreite 2 Jahre in  $t$ -Richtung und 1,33 Jahre in  $d$ -Richtung. Beide Achsen werden in Jahren gemessen. Die Werte von  $\hat{\alpha}$  werden als monatliche Hazards angegeben.

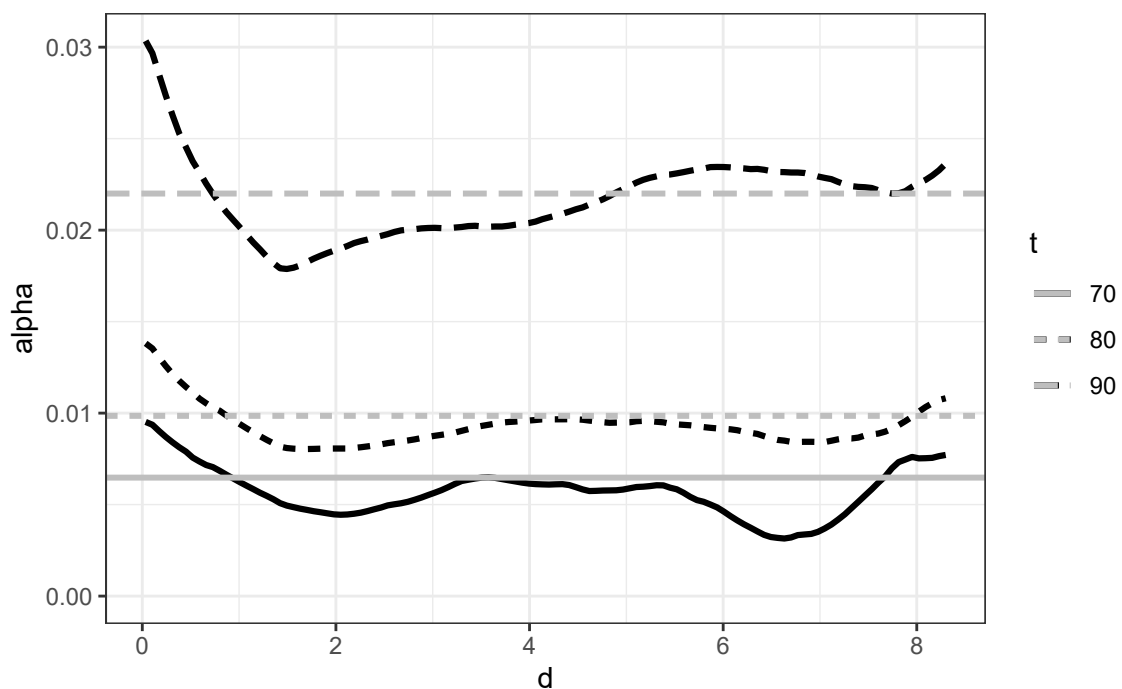


Abbildung 4.4: Drei Querschnitte für fixierte Alter  $t$  der Intensitätsfunktion. Außerdem sind Schätzungen unter der Markov-Annahme eingezeichnet (graue Linien), d.h. der Annahme, dass die Funktion  $\alpha$  nicht von  $d$  abhängt.

Abbildung 4.5 schließlich zeigt unterschiedliche Schätzungen von Hazardraten für ein einzelnes Individuum. Die unterste Kurve zeigt die geschätzte Sterbensintensität für eine Person, die nie eine Demenzdiagnose erhält, d.h. die Übergangsintensität vom Zustand „gesund“ in den Zustand „tot“. Die anderen beiden Kurven zeigen Sterbensintensitäten für eine Person, die im Alter von 75 Jahren eine Demenzdiagnose erhält, d.h.  $\alpha(t, t - 75)$  für  $t$  ab Diagnosezeitpunkt. Dies entspricht dem Querschnitt entlang einer Diagonalen in Abbildung 4.3. Eine Kurve zeigt die Schätzung unter der Markov-Annahme und eine ohne diese. Wie zu erwarten ist zu erkennen, dass von Demenz betroffene Personen unter deutlich erhöhtem Sterberisiko stehen. Der (relative) Unterschied zwischen den Kurven mit und ohne Markov-Annahme ist in den ersten Jahren nach Demenzdiagnose am größten, weshalb der rechte Teil der Abbildung die ersten fünf dieser Jahre hervorhebt.

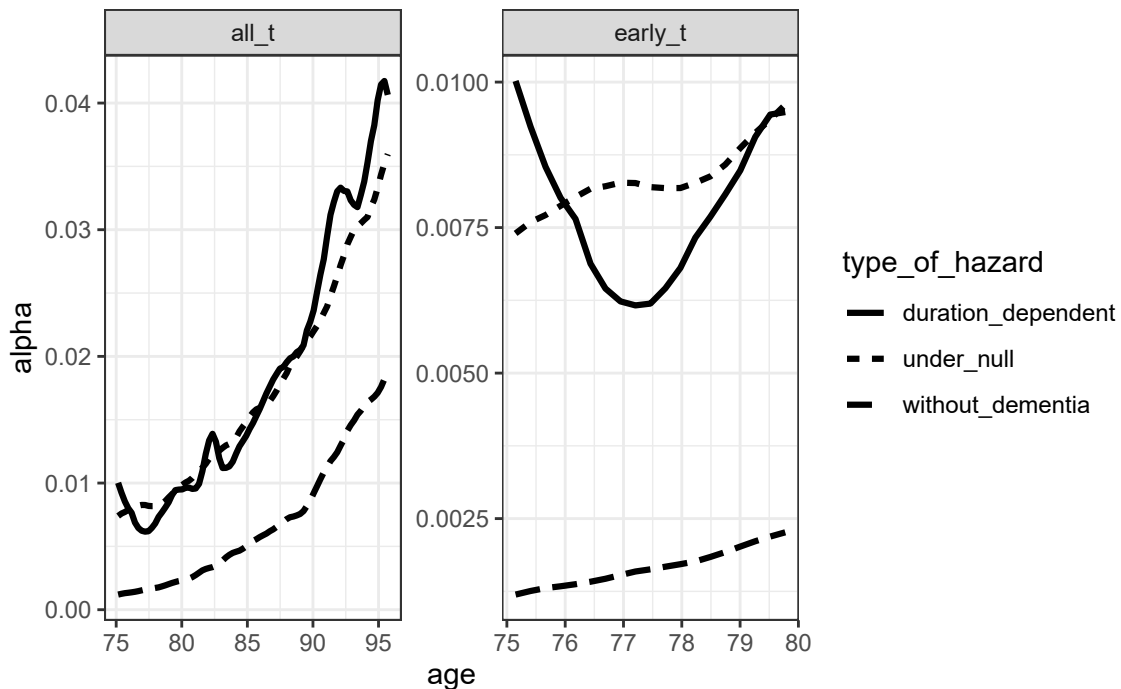


Abbildung 4.5: Sterbensintensität für Individuum ab Alter 75. Ohne Demenz (*without\_dementia*), bei Demenzdiagnose im Alter von 75 unter Markov-Annahme (*under\_null*) und ohne (*duration\_dependent*). Links: Für alle verfügbaren Alter ab 75. Rechts: Nur für die ersten fünf Jahre ab (möglicher) Diagnose.

## 4.4 Ergebnisse der statistischen Tests

Wir wollen nun die in Kapitel 2.4 vorgestellten Tests zur Anwendung bringen. Genauer gesagt wollen wir testen, ob die Funktion  $\alpha(t, d)$  von  $d$  abhängt. Dies ist gleichbedeutend mit der Frage, ob unser Modell markovsch ist. Wir beginnen mit dem Test, der ein festes Gitter nutzt (Gozalo-Typ I). Wir wählen dieses Gitter wie in Tabelle 4.1 dargestellt.

| $j$ | $t^j$ | $d^j$ |
|-----|-------|-------|
| 1   | 70    | 1     |
| 2   | 70    | 2     |
| 3   | 75    | 1     |
| 4   | 75    | 2     |
| 5   | 80    | 1     |
| 6   | 80    | 2     |

Tabelle 4.1: Gitter für Test vom Gozalo-Typ I. Achsen werden in Jahren gemessen.

Die angesichts des Datenumfangs kleine Gittergröße darf als konservativ angesehen werden, da so eher das Risiko besteht, dass kein Punkt dabei ist, an dem eine große Differenz der beiden Schätzungen vorliegt. Die Punkte stammen allesamt aus Bereichen der Funktion, in denen ausreichend Daten für eine gute Schätzung zu Verfügung stehen. Bei der Wahl der Punkte wurde die auf Seite 26 beschriebene Struktur aufgegriffen. Da das Thema der Bandbreitenwahl nicht Teil dieser Arbeit ist, wurde der Test für eine Reihe möglicher Bandbreiten durchgeführt, wobei das Verhältnis von  $t$ -Bandbreite zur  $d$ -Bandbreite sowie das Verhältnis von  $t$ -Bandbreite zur Bandbreite unter Nullhypothese konstant gelassen wurde (1,5 bzw. 1,3). Abbildung 4.6 zeigt die resultierenden p-Werte als Funktion der Bandbreite in  $t$ -Richtung. Die Werte liegen allesamt unter einem Signifikanzniveau von 5%, reichen im Maximum aber nahe an dieses heran. Die p-Werte hängen stark von der Wahl der Bandbreite ab, die binäre Testentscheidung zu einem Signifikanzniveau von 5% in diesem Fall nicht. Die größten p-Werte ergeben sich für  $t$ -Bandbreiten im Bereich zwischen acht und fünfzehn Monaten.

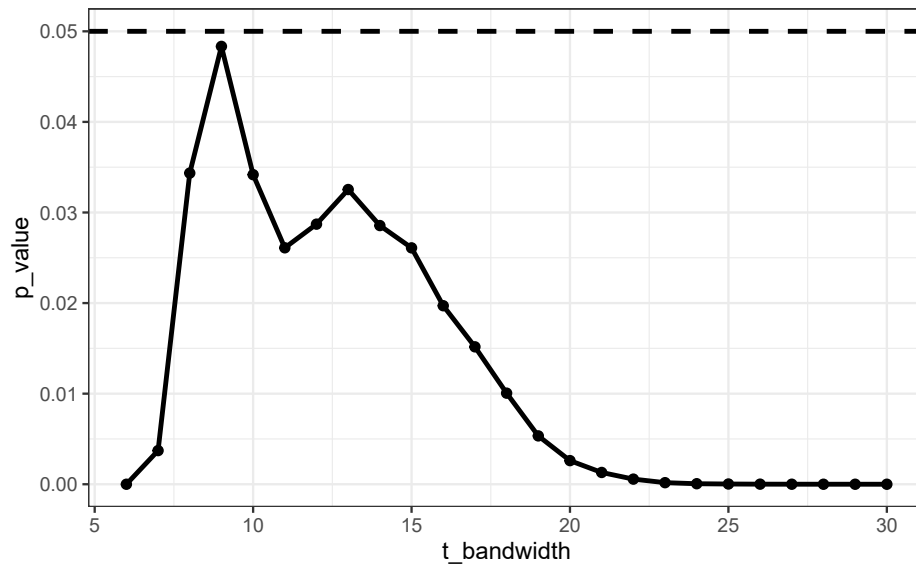


Abbildung 4.6: P-Werte für Test mit festem Gitter, abhängig von den genutzten Bandbreiten. Die Bandbreite in  $d$ -Richtung sowie die Bandbreite für die Schätzung unter Nullhypothese wurden im festen Verhältnis zur Bandbreite in  $t$ -Richtung gewählt. Die Bandbreite wird hier in Monaten gemessen. Außerdem ist ein Signifikanzniveau von 0,05 eingezeichnet.

Wir wollen außerdem einen Test vom Gozalo-Typ II durchführen. Als Größe des zufälligen Gitters wählen wir 15. Das Gitter wird durch u.i.v. Ziehungen aus der Gleichverteilung auf dem Rechteck  $[70, 90] \times [2, 4]$  ermittelt (Messung der Achsen in Jahren). Dieses Rechteck entstammt einem Bereich, in dem hinreichend viele Daten für gute Schätzungen vorliegen. Neben den Bandbreiten sind hier ebenfalls die Parameter  $\eta$  und  $\delta$  zu wählen. Abbildung 4.7 zeigt die resultierenden p-Werte als Funktion der Bandbreite in  $t$ -Richtung für drei unterschiedliche Kombinationen von  $\eta$  und  $\delta$ <sup>1</sup>. Es sei angemerkt, dass, da die Abhängigkeit der Hyperparameter im Fokus liegen soll, das selbe zufällig gewählte Gitter für alle Testausführungen genutzt wurde. Würde man jeweils ein neues zufälliges Gitter wählen, wären die Verläufe weniger glatt. Es fällt auf, dass die Testergebnisse stark von der Wahl von  $\eta$  und  $\delta$  abhängen. Für zwei der Kombinationen liegt die gesamte Kurve klar unter einem Signifikanzniveau von 5% (auch klarer als im Fall des Tests mit fixem Gitter), für die dritte Kombination kommen für gewisse Bandbreiten jedoch auch p-Werte von bis zu über 0.4 vor.

Abbildung 4.8 kann eine Idee geben, warum derartige Ergebnisse zu Stande kommen. Sie stellt sowohl den Strafterm  $\eta(n \cdot b_t \cdot b_d)^\delta$  (siehe Formel (2.7)) als auch geschätzte

<sup>1</sup>Die Fallzahl  $n$  fließt offensichtlich in den Strafterm  $\eta(n \cdot b_t \cdot b_d)^\delta$  (siehe Formel (2.7)) ein. Wie weiter oben erläutert ist das theoretisch korrekte  $n$  in unserer Anwendung die Anzahl der Frauen. In unserer Implementierung wurde allerdings die Anzahl der Frauen mit Demenzdiagnose genutzt. Dies ist insofern irrelevant, als dass nur die Wahl von  $\eta$  angepasst werden muss, um einen äquivalenten Test zu erhalten. Die hier genannten Werte von  $\eta$  beziehen sich auf den Fall, in dem  $n$  nur der Anzahl der an Demenz erkrankten Frauen entspricht.

Statistiken (insbesondere den Median und das 0.95-Quantil) der Verteilung von  $S_n^2(\tilde{X})$  dar (gegeben unsere Daten). Ist der Strafterm groß im Vergleich zu den größten Werten der Verteilung von  $S_n^2(\tilde{X})$ , so wird mit hoher Wahrscheinlichkeit die Teststatistik anhand der Stelle  $\tilde{X}^1$  ermittelt. Bei dieser einzelnen Stelle ist die Wahrscheinlichkeit relevant, dass nur eine so moderate Abweichung zwischen  $\hat{\alpha}$  und  $\hat{\beta}$  besteht, dass der Test keine Signifikanz anzeigt. Ist der Strafterm jedoch relativ gesehen klein, wird das maximale  $S_n^2(\tilde{X}^j)$  als Teststatistik gewählt. In diesem Fall ist die Abweichung dann mit hoher Wahrscheinlichkeit signifikant. Abbildung 4.8 illustriert, dass im Bereich von Bandbreiten zwischen acht und fünfzehn Monaten das Verhältnis zwischen den Straftermen und den größten Werten von  $S_n^2(x)$  am größten ist (daher die Hügel in diesem Bereich in Abbildung 4.7) und außerdem für  $\eta = 0,001$  und  $\delta = 0,6$  das Verhältnis besonders zu Gunsten des Strafterms ist (daher sieht man insignifikante Resultate für diese Wahl von  $\eta, \delta$ ).

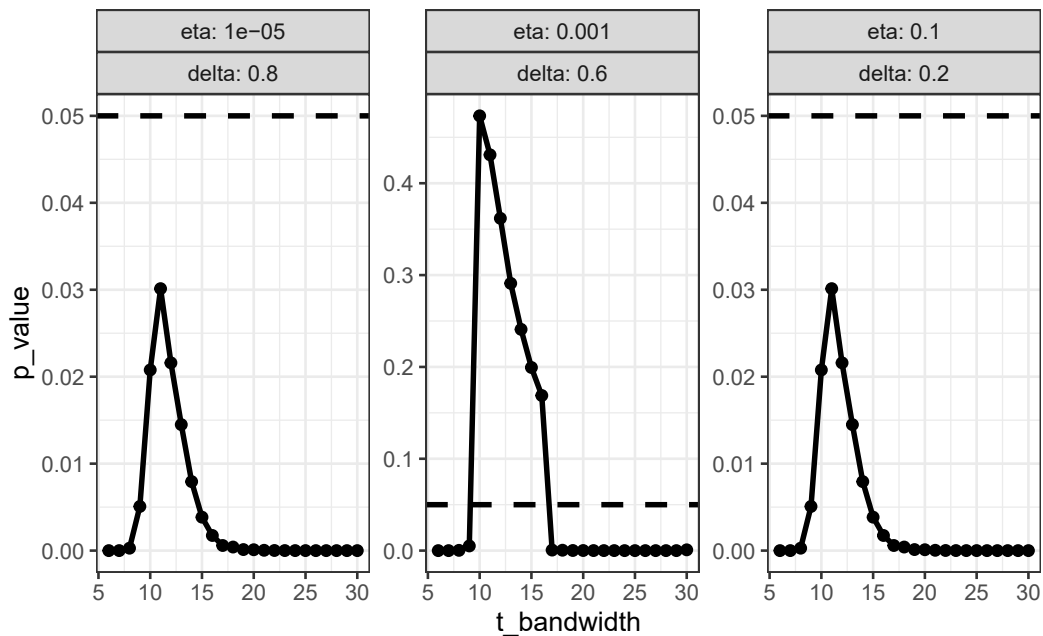


Abbildung 4.7: P-Werte für Test mit zufälligem Gitter, abhängig von den genutzten Bandbreiten. Die Bandbreite in  $d$ -Richtung sowie die Bandbreite für die Schätzung unter Nullhypothese wurden im festen Verhältnis zur Bandbreite in  $t$ -Richtung gewählt. Es werden die Resultate für drei unterschiedliche Kombinationen der Parameter  $\eta$  und  $\delta$  gezeigt. Da die Abhängigkeit der Ergebnisse von den Bandbreiten sowie  $\eta$  und  $\delta$  im Fokus steht, wurde jeder Test mit dem selben, zufällig gewählten Gitter durchgeführt. Außerdem ist jeweils ein Signifikanzniveau von 0,05 eingezeichnet.

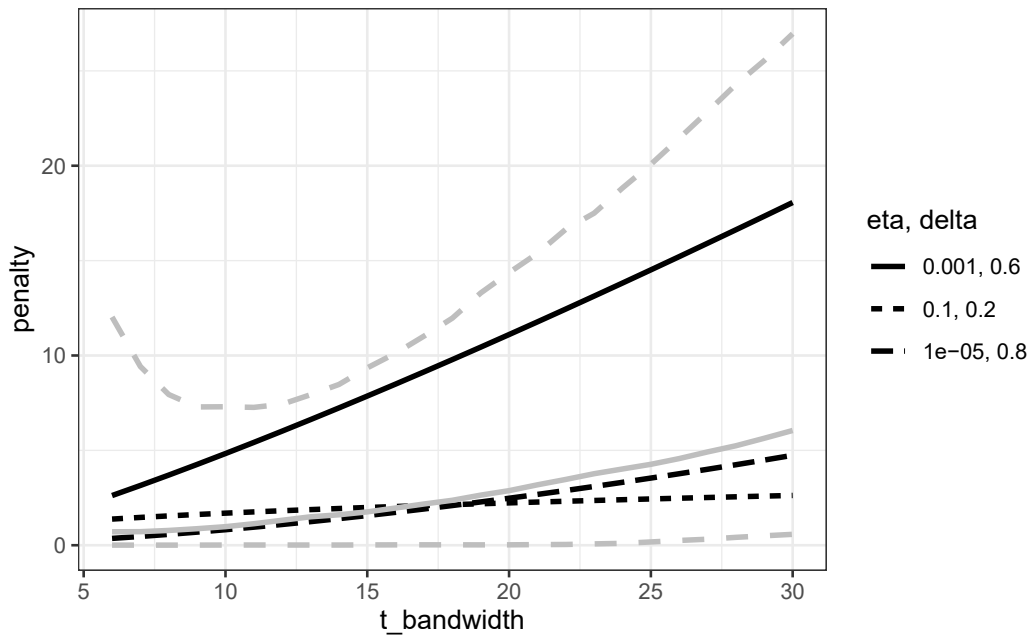


Abbildung 4.8: Der Plot zeigt für die drei Kombinationen von  $\eta$  und  $\delta$  und eine Reihe von Bandbreiten jeweils die Größe des Strafterms (schwarze Linien). Ergänzend dazu sind für alle Bandbreiten Schätzungen des 0,05-, 0,5- und 0,95-Quantils der Verteilung von  $S_n^2(\tilde{X})$  gegeben die Daten dargestellt (graue Linien).

## 4.5 Interpretation

Menschen können zu beliebigen Zeitpunkten in kontinuierlicher Zeit dement werden oder sterben und es erscheint klar, dass Sterbewahrscheinlichkeiten für demente Personen höher sind als für gleichaltrige Personen ohne derartige Erkrankung – dieses Geschehen wird durch unser Modell adäquat abgebildet. Unsere Daten hingegen weisen nicht diese detaillierte Auflösung auf. Zunächst liegen alle Zeitpunkte nur diskretisiert vor: Geburts- und Sterbedaten monatsgenau, Demenzdiagnosen quartalsgenau. Zum anderen sind die Daten zu Demenzdiagnosen genau das – Zeitpunkte der *Diagnosen*, nicht unbedingt des tatsächlichen Beginns der Krankheit. Besonders gut zu erkennen ist dies an den vielen Beobachtungen, bei denen Patienten sehr kurz nach ihrer Demenzdiagnose sterben, was sich in den obigen Schätzungen widerspiegelt im schnellen Abfall der Funktion bei kleinen  $d$ . Diese erklärt sich dadurch, dass häufig der Fall eintritt, dass schwerkranke Personen in ein Krankenhaus eingeliefert werden und dort noch kurz vor ihrem Tod eine Demenzdiagnose erhalten. In vielen Fällen litten diese Leute wahrscheinlich schon eine Weile unter einer Demenz, es hatte nur niemand die Diagnose gestellt (siehe etwa auch Huo u. a., 2018, die dieses Phänomen im Falle von Typ-2 Diabetes beobachten). Schließlich ist fraglich, inwieweit die Versicherten der AOK repräsentativ für die Gesamtbevölkerung sind. Der Horizont einer möglichen Inferenz über die Daten hinaus

ist also beschränkt. Diese Limitierungen der Daten müssen bei der Interpretation von Ergebnissen berücksichtigt werden.

Es fällt auf, dass die Formen der Funktion  $\hat{\alpha}(t, d)$  für verschiedene feste  $t$  jeweils in  $d$ -Richtung recht ähnlich sind, siehe Abbildungen 4.3 und 4.4. Dies könnte ein Anlass sein, um zu untersuchen, ob ein nichtparametrisches multiplikatives Modell (wie etwa in Linton u. a., 2003) geeignet wäre. Auch diese Frage könnte mit den von uns in Kapitel 2.4 beschriebenen Tests beantwortet werden. Um in diesen Rahmen eingeordnet werden zu können, muss nur die Menge  $\mathcal{B}$  der Nullhypothese entsprechend gewählt und ein geeigneter Schätzer für das multiplikative Modell berechnet werden. Angesichts des in  $t$ -Richtung annähernd exponentiellen Verlaufes von  $\hat{\alpha}$  wäre auch ein semi-parametrisches Cox-Modell der Form  $\alpha(t, d) = h(d) \exp(\beta \cdot t)$  denkbar. Dies könnte ebenfalls mit der selben Methodik getestet werden.



# Kapitel 5

## Diskussion

Das Problem, dass Prozesse nicht markovsch sind, kommt häufig in der Analyse mit Mehrzustandsmodellen auf. Wir präsentieren Methodik für eine konsistente nichtparametrische Intensitätsschätzung für eine gewisse Klasse an Modellen in diesem Fall. Darüber hinaus etablieren wir zwei Anpassungstests, welche zum Testen der Markov-Annahme genutzt werden können.

Es erscheint sehr plausibel, dass das Vorgehen in Kapitel 3.1 leicht auf allgemeine progressive Modelle verallgemeinerbar ist. Dabei handelt es sich grob gesagt um Mehrzustandsmodelle, bei denen in Strukturdiagrammen der Art von Abbildung 1.1 oder 1.2 in jeden Zustand nur ein Pfeil hinein zeigt, außer in den Anfangszustand, in welchen kein Pfeil hinein zeigen darf. In diesem Fall ist aus dem gegenwärtigen Zustand bereits herleitbar, welche Zustände der Prozess bereits besucht hat und in welcher Reihenfolge. Insofern ist die gesamte Historie des Prozesses aus dem gegenwärtigen Zustand sowie den bisherigen Sprungzeiten rekonstruierbar. Dementsprechend kann man Übergangintensitäten, die auf die gesamte Historie des Prozesses bedingen, als Intensitäten bedingt auf einen endlich-dimensionalen Zufallsvektor – den Verweildauern in den bisherigen Zuständen – ausdrücken. Diese Struktur ist wieder mit der Methodik nach NL schätzbar. Der Fluch der Dimension kann hier schließlich problematisch werden, da die Anzahl der Kovariablen offensichtlich ansteigt.

Möchte man nicht-markovsche Einflüsse außerhalb dieser beiden Klassen – progressive Modelle und verweildauerabhängige Modelle (wie der Semi-Markov-Fall) – modellieren, so sind allgemeine Ansätze wenig aussichtsreich, wie Hougaard (2000), S. 169f., diskutiert. Stattdessen muss situationsspezifisches Fachwissen genutzt werden, um zu berücksichtigen, auf welche Art und Weise die Prozesshistorie Information über die zukünftige Entwicklung enthält. Ein Beispiel für einen derartigen Versuch ist die Arbeit von dos Reis u. a. (2020), welche ein sogenanntes „Rating Momentum“ zu modellieren versucht.

In dieser Arbeit wurde nur ein Test auf Verweildauerabhängigkeit einer einzelnen

Übergangsintensität entwickelt. Interessant könnte aber darüber hinaus ein simultaner Test auf Verweildauerabhängigkeit von Übergangsintensitäten zwischen mehreren Paaren von Zuständen sein. Dafür müsste die Kovarianzstruktur der dazugehörigen Intensitätsschätzer untersucht werden. Im Falle der Unabhängigkeit könnte erneut eine  $\chi^2$ -Verteilung mit entsprechend mehr Freiheitsgraden zur Anwendung kommen.

Die in Kapitel 2 entwickelten Tests wurden in dieser Arbeit nur zum Test auf Verweildauerabhängigkeit einer Übergangsintensität genutzt. Wie bereits angedeutet ist der mögliche Anwendungsbereich allerdings deutlich weiter. Indem die Menge  $\mathcal{B}$  der Nullhypothese adäquat gewählt wird, könnte man ebenfalls auf die Validität bestimmter parametrischer, semiparametrischer oder strukturierter nichtparametrischer Modelle hin testen. Notwendig ist jeweils nur, dass ein unter der Nullhypothese geeigneter Schätzer bekannt ist, der in diesem Falle insbesondere schneller konvergiert als der Schätzer nach NL.

Das in der Verweildaueranalyse häufig präsente Problem der Rechtszensierung wurde in dieser Arbeit berücksichtigt. Es muss allerdings zugegeben werden, dass in einigen Fällen auch Linkszensierung vorhanden sein kann, (siehe etwa Weißbach u. Schmal, 2019). Daneben ist ohnehin in jeder Anwendung konkret zu überprüfen, inwieweit die hier gemachten Annahmen angemessen wirken. Zeitstetigkeit erscheint häufig natürlich, aber Daten entsprechen nicht immer dieser theoretischen Idee. Als Beispiel siehe die Situation der Daten aus Kapitel 4. Zensierung könnte informativ sein oder Beobachtungen nicht stochastisch unabhängig. Derartige Aspekte müssen jeweils im Lichte einer konkreten Situation erörtert werden.

Die Ergebnisse in Kapitel 4 verdeutlichen, dass eine starke Abhängigkeit der Resultate von der Wahl der Hyperparameter vorliegt. Dabei wurden einige derartiger Hyperparameter in unserer Analyse noch nicht diskutiert, sondern nur auf einen einzelnen Wert festgesetzt, etwa das Gitter für den Test von Typ I und die Gittergröße  $m_n$  und die Verteilung der Gitterpunkte beim Typ-II-Test. Insbesondere haben wir in letzterem Fall die Gittergröße für alle Bandbreiten konstant belassen. Die Annahme  $m_n = o((nb^{d+1})^\delta)$  (siehe (2.6)) suggeriert allerdings, dass die Gittergröße wachsend mit wachsenden Bandbreiten gewählt werden sollte.

Diese Arbeit befasst sich bewusst nicht mit systematischen Verfahren zur Auswahl von Bandbreiten und anderen Hyperparametern. Es existieren sowohl heuristische als auch datengetriebene Ansätze. NL nutzen etwa einen Leave-One-Out-Kreuzvalidierungsansatz, welcher von Vorschlägen in Nielsen (1990) und Andersen u. a. (1993) für den Fall ohne Kovariate inspiriert ist. Es sei darauf hingewiesen, dass für ein derartiges Verfahren bei unseren Daten die nötigen Rechenkapazitäten zu einem relevanten Thema werden.

# Literaturverzeichnis

- [Ait-Sahalia u. a. 2001] AIT-SAHALIA, Y. ; BICKEL, P.J. ; STOKER, T.M.: Goodness-of-fit tests for kernel regression with an application to option implied volatilities. In: *Journal of Econometrics* 105 (2001), S. 363–412
- [Andersen u. a. 1993] ANDERSEN, P. K. ; BORGAN, Ø; GILL, R. D. ; KEIDING, N.: *Statistical models based on counting processes*. Springer, New York, 1993
- [Andersen u. Gill 1982] ANDERSEN, P. K. ; GILL, R. D.: Cox’s regression model for counting processes: A large sample study. In: *The Annals of Statistics* 10 (1982), Nr. 4, S. 1100–1120
- [Andersen u. Keiding 2002] ANDERSEN, P. K. ; KEIDING, N.: Multi-state models for event history analysis. In: *Statistical Methods in Medical Research* 11 (2002), S. 91–115
- [Aspinall u. a. 2006] ASPINALL, W.P. ; CARNIEL, R. ; JAQUET, O. ; WOO, G. ; HINCKS, T.: Using hidden multi-state Markov models with multi-parameter volcanic data to provide empirical evidence for alert level decision-support. In: *Journal of Volcanology and Geothermal Research* 153 (2006), S. 112–124
- [Bao u. a. 2019] BAO, M. ; DING, Y. ; SINGH, C. ; SHAO, C.: A Multi-State Model for Reliability Assessment of Integrated Gas and Power Systems Utilizing Universal Generating Function Techniques. In: *IEEE Transactions on Smart Grid* 10 (2019), S. 6271–6283
- [Bauer 1992] BAUER, Heinz: *Maß- und Integrationstheorie*. de Gruyter, Berlin, 1992
- [Bierens 1984] BIERENS, H.J.: Model specification testing of time series regressions. In: *Journal of Econometrics* 26 (1984), S. 323–353
- [Biessy 2017] BIESSY, G.: Continuous-time Semi-Markov inference of biometric laws associated with a long-term care insurance portfolio. In: *ASTIN Bulletin: The Journal of the IAA* 47 (2017), S. 527–561
- [Billingsley 2012] BILLINGSLEY, Patrick: *Probability and Measure*. Wiley, New York, 2012

- [Bluhm u. a. 2002] BLUHM, C. ; OVERBECK, L. ; WAGNER, C.: *An Introduction to Credit Risk Modeling*. Chapman & Hall, Boca Raton, 2002
- [Chung 2001] CHUNG, Kai L.: *A Course in Probability Theory*. Academic Press, San Diego, 2001
- [Commenges 1999] COMMENGES, D.: Multi-state Models in Epidemiology. In: *Lifetime Data Analysis* 5 (1999), S. 315–327
- [Corraini u. a. 2017] CORRAINI, P. ; HENDERSON, V.W. ; ORDING, A.G. ; PEDERSEN, L. ; HORVÁTH-PUHÓ, E. ; SØRENSEN, H.T.: Long-Term Risk of Dementia Among Survivors of Ischemic or Hemorrhagic Stroke. In: *Stroke* 48 (2017), S. 180–186
- [Cox 1972] COX, D. R.: Regression models and life tables. In: *Journal of the Royal Statistical Society* 34 (1972), Nr. 2, S. 187–220
- [D’Amico u. a. 2017] D’AMICO, G. ; BIASE, G. D. ; JANSSEN, J. ; MANCA, R.: *Semi-Markov migration models for credit risk*. Wiley, Hoboken, 2017
- [Datta u. Satten 2001] DATTA, Somnath ; SATTEN, Glen A.: Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. In: *Statistics & Probability Letters* 55 (2001), Nr. 4, S. 403–411
- [Devolder u. a. 2015] DEVOLDER, P. ; JANSSEN, J. ; MANCA, R.: *Basic Stochastic Processes*. Wiley, Hoboken, 2015
- [Elstrodt 2009] ELSTRODT, Jürgen: *Maß- und Integrationstheorie*. Springer, Berlin, 2009
- [Fan u. Li 1996] FAN, Y. ; LI, Q.: Consistent model specification tests: Omitted variables and semiparametric functional forms. In: *Econometrica* 64 (1996), Nr. 4, S. 865–890
- [Fleming u. Harrington 2011] FLEMING, Thomas R. ; HARRINGTON, David P.: *Counting processes and survival analysis*. Wiley, Hoboken, 2011
- [Gozalo 1993] GOZALO, Pedro L.: A consistent model specification test for nonparametric estimation of regression function models. In: *Econometric Theory* 9 (1993), S. 451–477
- [Grabski 2015] GRABSKI, F.: *Semi-Markov processes: applications in system reliability and maintenance*. Elsevier, Amsterdam, 2015
- [Hjort 1994] HJORT, N. L.: Dynamic likelihood hazard rate estimation. In: *Manuscript* (1994)

- [Hougaard 1999] HOUGAARD, Philip: Multi-state Models: A Review. In: *Lifetime Data Analysis* 5 (1999), S. 239–264
- [Hougaard 2000] HOUGAARD, Philip: *Analysis of Multivariate Survival Data*. Springer, New York, 2000
- [Huo u. a. 2018] HUO, L. ; MAGLIANO, D. J. ; RANCI‘ERE, F. ; HARDING, J. L. ; NANAYAKKARA, N. ; SHAW, J. E. ; CARSTENSEN, B.: Impact of age at diagnosis and duration of type 2 diabetes on mortality in Australia 1997–2011. In: *Diabetologia* 61 (2018), S. 239–264
- [Janssen u. Manca 2006] JANSSEN, J. ; MANCA, R.: *Applied Semi-Markov Processes*. Springer, New York, 2006
- [Jones u. a. 1994] JONES, M. C. ; DAVIES, S. J. ; PARK, B. U.: Versions of kernel-type regression estimators. In: *Journal of the American Statistical Association* 89 (1994), S. 825–832
- [Keiding 1986] KEIDING, N.: Statistical analysis of semi-Markov processes based on the theory of counting processes. In: J., Janssen (Hrsg.): *Semi-Markov Models*. Springer, 1986, S. 301–315
- [Lando u. Skødeberg 2002] LANDO, David ; SKØDEBERG, Torben M.: Analyzing rating transitions and rating drift with continuous observations. In: *Journal of Banking & Finance* 26 (2002), Nr. 2-3, S. 423–444
- [Lavergne u. a. 2015] LAVERGNE, P. ; MAISTRE, S. ; PATILEA, V.: A significance test for covariates in nonparametric regression. In: *Electronic Journal of Statistics* 9 (2015), S. 643–678
- [Lavergne u. Vuong 2000] LAVERGNE, P. ; VUONG, Q.: Nonparametric significance testing. In: *Econometric Theory* 16 (2000), S. 576–601
- [Linton u. a. 2003] LINTON, Oliver B. ; NIELSEN, Jens P. ; GEER, Sara van d.: Estimating multiplicative and additive hazard functions by kernel methods. In: *The Annals of Statistics* 31 (2003), Nr. 2, S. 464–492
- [McKeague u. Utikal 1990a] MCKEAGUE, I. W. ; UTIKAL, K. J.: Identifying nonlinear covariate effects in semimartingale regression models. In: *Probability Theory and Related Fields* 87 (1990), S. 1–25
- [McKeague u. Utikal 1990b] MCKEAGUE, I. W. ; UTIKAL, K. J.: Inference for a nonlinear counting process regression model. In: *The Annals of Statistics* 18 (1990), S. 1172–1187

- [Nadaraya 1989] NADARAYA, E. A.: *Nonparametric estimation of probability densities and regression curves*. Springer Netherlands, Dordrecht, 1989
- [Nielsen u. Linton 1995] NIELSEN, Jens P. ; LINTON, Oliver B.: Kernel Estimation in a Nonparametric Marker Dependent Hazard Model. In: *The Annals of Statistics* 23 (1995), Nr. 5, S. 1735–1748
- [Nielsen 1990] NIELSEN, J.P.: *Kernel estimation of densities and hazards: a counting process approach*, Univ. California, Berkeley, Diss., 1990
- [Pendlebury u. Rothwell 2009] PENDLEBURY, S.T. ; ROTHWELL, P.M.: Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: A systematic review and meta-analysis. In: *The Lancet Neurology* 8 (2009), S. 1006–1018
- [PROVA-Study-Group 1991] PROVA-STUDY-GROUP: Prophylaxis of first hemorrhage from esophageal varices by sclerotherapy, propranolol or both in cirrhotic patients: A randomized multicenter trial. In: *Hepatology* 14 (1991), S. 1016–1024
- [Ramlaau-Hansen 1983] RAMLAU-HANSEN, H.: Smoothing counting process intensities by means of kernel functions. In: *Annals of Statistics* 11 (1983), S. 453–466
- [Rebolledo 1980] REBOLLEDO, R.: Central limit theorems for local martingales. In: *Zeitschrift für Wahrscheinlichkeitstheorie* 51 (1980), S. 269–286
- [dos Reis u. a. 2020] REIS, G. dos ; PFEUFFER, M. ; SMITH, G.: Capturing model risk and rating momentum in the estimation of probabilities of default and credit rating migrations. In: *Quantitative Finance* 20 (2020), Nr. 7, S. 1069–1083
- [Titman 2015] TITMAN, Andrew C.: Transition probability estimates for non-Markov multi-state models. In: *Biometrics* 71 (2015), Nr. 4, S. 1034–1041
- [Una-Alvarez u. Meira-Machado 2015] UNA-ALVAREZ, Jacobo d. ; MEIRA-MACHADO, Luis: Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. In: *Biometrics* 71 (2015), Nr. 2, S. 364–375
- [van der Vaart 1998] VAART, A. W. d.: *Asymptotic Statistics*. Cambridge University Press, New York, 1998
- [Vital u. Patil 2020] VITAL, R.-A. ; PATIL, Prakash: Goodness-of-fit test for hazard rate. In: *Journal of Nonparametric Statistics* 32 (2020), S. 403–427
- [Weißbach u. a. 2021] WEISSBACH, R. ; KIM, Y. ; DÖRRE, A. ; FINK, A. ; DOBLHAMMER, G.: Left-Censored Dementia Incidences in Estimating Cohort Effects. In: *Lifetime Data Analysis* 27 (2021), S. 38–63

- [Weißbach u. Mollenhauer 2011] WEISSBACH, R. ; MOLLENHAUER, T.: Modeling rating transitions. In: *Journal of the Korean Statistical Society* 40 (2011), S. 469–485
- [Weißbach u. Schmal 2019] WEISSBACH, R. ; SCHMAL, F.: Estimating the probability of a non-Markovian rating transition from partially unobserved histories. In: *Journal of Risk Management in Financial Institutions* 12 (2019), S. 256–267
- [Witting 1985] WITTING, Hermann: *Mathematische Statistik I*. B. G. Teubner, Stuttgart, 1985

# Anhang A

## R-Code

### Transformation der Daten

```
library(tidyverse)
library(haven)
library(lubridate)

dementia_raw <- read_dta("../Data\\dementia_original.dta")

## summarize data into one line per person
dementia_per_person <- dementia_raw %>%
  group_by(VERSID) %>%
  summarize(year_of_censoring = ifelse(sum(tod) == 0, JAHR[n()], NA),
            quarter_of_censoring = ifelse(sum(tod) == 0, QUARTAL[n()], NA),
            year_of_dementia = ifelse(max(ever_flag_sicher_dem) == 1,
                                      JAHR[which(inci_dem == 1)], NA),
            quarter_of_dementia = ifelse(sum(ever_flag_sicher_dem) > 0,
                                      QUARTAL[which(inci_dem == 1)], NA),
            year_of_birth = GEBJAHR[1],
            month_of_birth = GEBMONAT[1],
            year_of_death = ifelse(sum(tod) > 0, TODJAHR[n()], NA),
            month_of_death = ifelse(sum(tod) > 0, TODMONAT[n()], NA),
            sex = if_else(GESCHL[1] == 1, "male", "female"))

## Three helper functions for transformation
mean_month_of_quarter <- function(quarter) {
  return((quarter - 1) * 3 + 1.5)
}
quarter_of_month <- function(month) {
  return(ceiling(month / 3))
}

month_of_dementia <- function(year_of_dementia, quarter_of_dementia,
                             year_of_death, month_of_death) {
  case_when(
    is.na(year_of_death) ~ mean_month_of_quarter(quarter_of_dementia),
    year_of_dementia < year_of_death ~ mean_month_of_quarter(quarter_of_dementia),
    ! quarter_of_month(month_of_death) == quarter_of_dementia ~
      mean_month_of_quarter(quarter_of_dementia),
    TRUE ~ (quarter_of_dementia - 1) * 3/2 + month_of_death/2
  )
}
```



```

## add correct months of censoring and dementia
dementia_per_person <- dementia_per_person %>%
  mutate(month_of_censoring = 3 * quarter_of_censoring ,
          month_of_dementia = month_of_dementia(year_of_dementia ,
                                                  quarter_of_dementia ,
                                                  year_of_death ,
                                                  month_of_death)) %>%

  rename(id = VERSID)

## another helper function for transformation
ym_diff <- function(year_new, month_new, year_old, month_old, return_months = T){
  if_else(
    rep(return_months, length(year_new)),
    12 * year_new + month_new - 12 * year_old - month_old ,
    (12 * year_new + month_new - 12 * year_old - month_old) / 12
  )
}

# dataframe that contains the critical values for analysis
dementia_analysis <- dementia_per_person %>%
  mutate(time_to_death = ym_diff(year_of_death, month_of_death,
                                year_of_birth, month_of_birth),
         time_to_dementia = ym_diff(year_of_dementia, month_of_dementia,
                                    year_of_birth, month_of_birth),
         time_to_censoring = ym_diff(year_of_censoring, month_of_censoring,
                                     year_of_birth, month_of_birth),
         birth = year_of_birth + month_of_birth / 12) %>%
  select(id, birth, time_to_death, time_to_dementia, time_to_censoring, sex)

write_csv(dementia_per_person, "..\\Data\\dementia_per_person.csv")

write_csv(dementia_analysis, "..\\Data\\dementia_analysis.csv")

```

## Funktionen für die Schätzung der Intensitäten

```

library(tidyverse)

source("nielsen_linton_tools_support.R")

#' compute Nielsen_Linton intensity estimate alpha(t,d)
#' for single pair of t and d and single pair of bandwidths
#' given a proper dataset, using UNIFORM kernel
#' ARGS
#'   data: dataframe with with 3 columns, named (exactly) Z1, Z2, DELTA
#'   t: real number, positive. t-coordinate where alpha(t,d) is to be estimated
#'   d: real number, positive. d-coordinate where alpha(t,d) is to be estimated
#'   needs to be less or equal to t
#'   t_bandwidth: real number, positive. bandwidth parameter in t direction
#'   d_bandwidth: real number, positive. bandwidth parameter in d direction
#' RETURN
#'   real number. estimate of intensity alpha(t,d), given data and bandwidths
pointwise_nl_estimate_uniform_kernel <- function(data, t, d,
                                                t_bandwidth, d_bandwidth) {
  if (d > t) stop("'d' can't be greater than 't'.")

  data %>% summarize(
    numerator = sum( (abs(Z2 - t) <= t_bandwidth) *
                    (abs(Z2 - Z1 - d) <= d_bandwidth) * (DELTA == 2) ),
    denominator = sum( pos( pmin(Z2, t + t_bandwidth, d + Z1 + d_bandwidth) -

```

```

      pmax(Z1, t - t_bandwidth, d + Z1 - d_bandwidth) ) ),
    alpha = numerator / denominator
  ) %>%
  select(alpha) %>%
  pull()
}

#' compute Nielsen-Linton intensity estimate alpha(t,d)
#' for a grid of ts, ds, and bandwidths using uniform kernel
#' ARGS
#'   data: dataframe with 3 columns, named (exactly) Z1, Z2, DELTA
#'   ts: vector of pos. numbers. t-coordinates where alpha is to be estimated
#'   ds: vector of pos. numbers. d-coordinates where alpha is to be estimated
#'   t_bws: vector of bandwidths ought to be used in t-direction
#'   d_bws: vector of bandwidths ought to be used in d-direction.
#'   Must be same length as t_bws
#' RETURN
#'   dataframe with 5 columns (t, d, t_bw, d_bw, alpha)
#'   and (length(ts) x length(ds) x length(t_bws)) rows
grid_nl_estimate_uniform_kernel <- function(data, ts, ds, t_bws, d_bws) {
  bws <- cbind(t_bws, d_bws)

  td_grid <- merge(ts, ds, by = NULL)
  grid_df <- merge(td_grid, bws, by = NULL) %>%
    as_tibble() %>%
    rename(t = x, d = y, t_bandwidth = t_bws, d_bandwidth = d_bws) %>%
    filter(d <= t)

  grid_df %>%
    mutate(
      alpha = pmap_dbl(
        .l = list(t, d, t_bandwidth, d_bandwidth),
        .f = ~ pointwise_nl_estimate_uniform_kernel(data, ..1, ..2, ..3, ..4)
      )
    )
}

#' Compute the denominator for Nielsen-Linton intensity estimate
#' can be recycled for variance estimate
#' ARGS
#'   Z1, Z2: same-length vectors of pos. numbers. Data the estimate is based on.
#'   t, d: real numbers. arguments, for which the denominator
#'         of the estimator ought to be evaluated
#'   t_bandwidth, d_bandwidth: real numbers, bandwidths in t and d direction.
#'   kernel: function with support [-1,1]
nl_denominator <- function(Z1, Z2, t, d, t_bandwidth, d_bandwidth, kernel) {
  lower_int_limit <- t - t_bandwidth
  upper_int_limit <- t + t_bandwidth

  denominator_integrand <- function(s) {
    map_dbl(s,
      ~sum(
        (Z1 < . & . <= Z2) * kernel((t - .) / t_bandwidth) *
        kernel((d - (. - Z1)) / d_bandwidth)
      )
    )
  }
}

return(

```

```

    integrate(
      f = denominator_integrand,
      lower = lower_int_limit,
      upper = upper_int_limit,
      rel.tol = .Machine$double.eps^0.05
    )$value
  )
}

#' As nl_denominator, just for the nullhypothesis-case, i.e. alpha not
#' depending on d. Hence, d and d_bandwidth are not required anymore.
nl_denominator_under_null <- function(Z1, Z2, t, bandwidth, kernel) {
  lower_int_limit <- t - bandwidth
  upper_int_limit <- t + bandwidth

  denominator_integrand <- function(s) {
    map_dbl(s,
      ~sum(
        (Z1 < . & . <= Z2) * kernel((t - .) / bandwidth)
      )
    )
  }

  return(
    integrate(
      f = denominator_integrand,
      lower = lower_int_limit,
      upper = upper_int_limit,
      #subdivisions = no_subdivisions,
      rel.tol = .Machine$double.eps^0.05
    )$value
  )
}

#' Similar to nl_denominator and nl_denominator_under_null,
#' just for the no-covariate-case.
#' Here the data required are one-dimensional.
hjort_denominator <- function(Z, t, bandwidth, kernel) {
  lower_int_limit <- t - bandwidth
  upper_int_limit <- t + bandwidth

  denominator_integrand <- function(s) {
    map_dbl(s,
      ~sum(
        (. <= Z) * kernel((t - .) / bandwidth)
      )
    )
  }

  return(
    integrate(
      f = denominator_integrand,
      lower = lower_int_limit,
      upper = upper_int_limit,
      #subdivisions = no_subdivisions,
      rel.tol = .Machine$double.eps^0.05
    )$value
  )
}

```

```

#' compute Nielsen_Linton intensity estimate alpha(t,d)
#' for single pair of t and d and single pair of bandwidths
#' given a proper dataset, using arbitrary kernel with bounded support
#' integral in the denominator is approximated numerically
#' ARGS
#'   data: dataframe with with 3 columns, named (exactly) Z1, Z2, DELTA
#'   t: real number, positive. t-coordinate where alpha(t,d) is to be estimated
#'   d: real number, positive. d-coordinate where alpha(t,d) is to be estimated
#'       needs to be less or equal to t
#'   t_bandwidth: pos. real number. bandwidth (hyper-)parameter in t direction
#'   d_bandwidth: pos. real number. bandwidth (hyper-)parameter in d direction
#'   kernel: function. kernel function with support [-1,1]
#' RETURN
#'   real number. estimate of intensity alpha(t,d), given data and bandwidths
pointwise_nl_estimate <- function(data, t, d, t_bandwidth, d_bandwidth, kernel) {
  if (d > t) stop("'d' can't be greater than 't'.")

  Z1 <- data$Z1
  Z2 <- data$Z2
  DELTA <- data$DELTA

  numerator <- sum(
    (DELTA == 2) * kernel((t - Z2) / t_bandwidth) *
    kernel((d - (Z2 - Z1)) / d_bandwidth)
  )
  denominator <- nl_denominator(Z1, Z2, t, d, t_bandwidth, d_bandwidth, kernel)

  return(numerator / denominator)
}

#' compute Nielsen_Linton intensity estimate alpha(t,d) for a grid of ts, ds,
#' and bandwidths, given a bounded kernel
#' ARGS
#'   data: dataframe with with 3 columns, named (exactly) Z1, Z2, DELTA
#'   ts: vector of pos. numbers. t-coordinates where alpha is to be estimated
#'   ds: vector of pos. numbers. d-coordinates where alpha is to be estimated
#'   t_bws: vector of bandwidths ought to be used in t-direction
#'   d_bws: vector of bandwidths ought to be used in d-direction.
#'       Must be same length as t_bws
#'   kernel: function. kernel function with support [-1,1]
#' RETURN
#'   dataframe with 5 columns(t, d, t_bw, d_bw, alpha) and
#'   (length(ts) x length(ds) x length(t_bws)) rows
grid_nl_estimate <- function(data, ts, ds, t_bws, d_bws, kernel) {
  bws <- cbind(t_bws, d_bws)

  td_grid <- merge(ts, ds, by = NULL)
  grid_df <- merge(td_grid, bws, by = NULL) %>%
    as_tibble() %>%
    rename(t = x, d = y, t_bandwidth = t_bws, d_bandwidth = d_bws) %>%
    filter(d <= t)

  grid_df %>%
    mutate(
      alpha = pmap_dbl(
        .l = list(t, d, t_bandwidth, d_bandwidth),
        .f = ~ pointwise_nl_estimate(data, ..1, ..2, ..3, ..4, kernel = kernel)
      )
    )

```

```

    )
  }

#' compute variance estimate of Nielsen_Linton intensity estimate
#' for single pair of t and d and single pair of bandwidths
#' given a proper dataset, using arbitrary kernel with bounded support
#' integral in the denominator is approximated numerically
#' ARGS
#'   data: dataframe with with 3 columns, named (exactly) Z1, Z2, DELTA
#'   t: pos. real number. t-coordinate where alpha(t,d) ought to be estimated
#'   d: pos. real number. d-coordinate where alpha(t,d) ought to be estimated
#'       needs to be less or equal to t
#'   t_bandwidth: real number, positive. bandwidth parameter in t direction
#'   d_bandwidth: real number, positive. bandwidth parameter in d direction
#'   kernel: function. kernel function with support [-1,1]
#' RETURN
#'   real number. estimate of variance of NL-estimate given data and bandwidths
pointwise_nl_variance_estimate <- function(data, t, d,
                                           t_bandwidth, d_bandwidth, kernel) {
  n <- nrow(data)
  Z1 <- data$Z1
  Z2 <- data$Z2
  DELTA <- data$DELTA

  numerator <- sum(
    (DELTA == 2) * kernel((t - Z2) / t_bandwidth)^2
    * kernel((d - (Z2 - Z1)) / d_bandwidth)^2
  )
  denominator <- nl_denominator(Z1, Z2, t, d, t_bandwidth, d_bandwidth, kernel)^2

  return(n * t_bandwidth * d_bandwidth * numerator / denominator)
}

#' similar to pointwise_nl_estimate, just for the nullhypothesis-case of
#' alpha not depending on d. Hence no arguments d, d_bandwidth required
pointwise_nl_estimate_under_null <- function(data, t, bandwidth, kernel) {
  Z1 <- data$Z1
  Z2 <- data$Z2
  DELTA <- data$DELTA

  numerator <- sum(
    (DELTA == 2) * kernel((t - Z2) / bandwidth)
  )
  denominator <- nl_denominator_under_null(Z1, Z2, t, bandwidth, kernel)

  return(numerator / denominator)
}

#' similar to grid_nl_estimate, just for the nullhypothesis-case of
#' alpha not depending on d. Hence no arguments ds, d_bws required
grid_nl_estimate_under_null <- function(data, ts, bws, kernel) {

  grid_df <- merge(ts, bws, by = NULL) %>%
    as_tibble() %>%
    rename(t = x, bandwidth = y)

  grid_df %>%
    mutate(
      alpha = pmap_dbl(

```

```

      .l = list(t, bandwidth),
      .f = ~ pointwise_nl_estimate_under_null(data, ..1, ..2, kernel = kernel)
    )
  )
}

#' Similar to pointwise_nl_estimate, just for the case without covariates
#' Hence, the structure of data is different
#' ARGS
#' data: dataframe with columns Z and DELTA; Z = min(T, U), DELTA = (T <= U)
pointwise_hjort_estimate <- function(data, t, bandwidth, kernel) {
  Z <- data$Z
  DELTA <- data$DELTA

  numerator <- sum(
    (DELTA == 1) * kernel((t - Z) / bandwidth)
  )
  denominator <- hjort_denominator(Z, t, bandwidth, kernel)

  return(numerator / denominator)
}

#' Similar to grid_nl_estimate, just for the case without covariates
#' Hence, the structure of data is different
#' ARGS
#' data: dataframe with columns Z and DELTA; Z = min(T, U), DELTA = (T <= U)
grid_hjort_estimate <- function(data, ts, bws, kernel) {

  grid_df <- merge(ts, bws, by = NULL) %>%
    as_tibble() %>%
    rename(t = x, bandwidth = y)

  grid_df %>%
    mutate(
      alpha = pmap_dbl(
        .l = list(t, bandwidth),
        .f = ~ pointwise_hjort_estimate(data, ..1, ..2, kernel = kernel)
      )
    )
}

```

## Funktionen für die Anpassungstests

```

#' Computes test statistics  $S_n^2$  for each point of a grid of points
#' given data, bandwidths and kernel function
#' ARGS
#' data: dataframe, with columns named Z1, Z2, DELTA
#' grid: dataframe with two numerical columns, named t and d.
#' grid points, on which test is based.
#' t_bandwidth, d_bandwidth, null_t_bandwidth: real numbers.
#' different bandwidths to use for estimation in context of the test
#' kernel: function. kernel function with support  $[-1,1]$ 
#' RETURN
#' numeric vector of length equal to nrow(grid)
grid_gozalo_test_statistics <- function(data, grid, t_bandwidth, d_bandwidth,
                                       null_t_bandwidth, kernel) {
  n <- nrow(data)
  p <- nrow(grid)

```

```

test_df <- grid %>%
  mutate(
    alternative_estimate = map2_dbl(.x = t, .y = d,
                                   ~pointwise_nl_estimate(data, t = .x,
                                                           d = .y,
                                                           t_bandwidth,
                                                           d_bandwidth,
                                                           kernel)),
    null_estimate = map_dbl(t, ~pointwise_nl_estimate_under_null(data, .,
                                                                  null_t_bandwidth,
                                                                  kernel)),
    variance_estimate = map2_dbl(.x = t, .y = d,
                                 ~pointwise_nl_variance_estimate(data,
                                                                    t = .x,
                                                                    d = .y,
                                                                    t_bandwidth,
                                                                    d_bandwidth,
                                                                    kernel)),
    std_error = sqrt( variance_estimate / (n * t_bandwidth * d_bandwidth) ),
    S_n_squared = ((alternative_estimate - null_estimate) / std_error )^2
  )
return(test_df$S_n_squared)
}

#' performs a Gozalo-type I test, given data, a grid of points,
#' bandwidths and a kernel function
#' ARGS
#'   data: dataframe, with columns named Z1, Z2, DELTA
#'   grid: dataframe with two numerical columns, named t and d. grid points,
#'         on which test is based.
#'   t_bandwidth, d_bandwidth, null_t_bandwidth: real numbers. different
#'         bandwidths to use for estimation in context of the test
#'   kernel: function. kernel function with support [-1,1]
#' RETURN
#'   numeric value between 0 and 1. p value for the test.
fixed_grid_gozalo_test <- function(data, grid, t_bandwidth, d_bandwidth,
                                   null_t_bandwidth, kernel) {

  test_statistic <- sum(grid_gozalo_test_statistics(data, grid, t_bandwidth,
                                                    d_bandwidth, null_t_bandwidth, kernel))

  p_value <- pchisq(test_statistic, lower.tail = FALSE, df = nrow(grid))

  return(p_value)
}

#' gives grid of uniformly distributed points from some rectangle
#' in format suited for gozalo test functions
#' ARGS
#'   t_limits, d_limits: numeric vectors of length 2 each.
#'   limits of rectangle to choose grid from
#'   p: number of grid points
#' RETURN
#'   dataframe with two numerical columns, named t and d
uniformly_random_grid <- function(t_limits, d_limits, p) {
  tibble(
    t = runif(p),
    d = runif(p)
  ) %>%

```

[illegible]



```

        .f = ~grid_search_gozalo_test(
          data = data,
          grid = random_grid,
          t_bandwidth = ..1,
          d_bandwidth = ..2,
          null_t_bandwidth = ..3,
          kernel = kernel,
          eta = ..4,
          delta = ..5)
      )
    )
  } else if (point_grid == "each_new") {
    hyperparameter_grid %>%
      mutate(p_value = pmap_dbl(.l = list(t_bandwidth, d_bandwidth,
                                           null_t_bandwidth, eta, delta),
                                .f = ~grid_search_gozalo_test(
                                  data = data,
                                  grid = uniformly_random_grid(t_limits,
                                                                d_limits,
                                                                no_grid_points),
                                  t_bandwidth = ..1,
                                  d_bandwidth = ..2,
                                  null_t_bandwidth = ..3,
                                  kernel = kernel,
                                  eta = ..4,
                                  delta = ..5)
                                )
            )
  }
}

```

## Wahl der Hyperparameter

```

# The choices of hyperparameters for data analysis in the context of my thesis
# are concentrated in this script.

```

```

# ratios between different kind of badnwidths
# these are fixed in order to vary bandwidth only in one dimension
t_to_d_bandwidth_ratio <- 1.5
t_to_null_t_bandwidth_ratio <- 1.3

```

```

kernel <- triangular_kernel

```

```

## bandwidths for NL estimation
t_bandwidths <- 6:30
d_bandwidths <- t_bandwidths / t_to_d_bandwidth_ratio

```

```

## bandwidths for NL estimation under null
null_t_bandwidths <- t_bandwidths / t_to_null_t_bandwidth_ratio

```

```

## bandwidths for Hjort estimation for non-dementia hazard-of-death rates
# note: due to much more data, bandwidth can probably be set lower here
hjort_t_bandwidths <- 5:25

```

```

## hyperparameters for fixed grid Gozalo tests
grid_ts <- 12 * c(70, 70, 75, 75, 80, 80)
grid_ds <- 12 * rep(c(2, 4), 3)
fixed_grid <- tibble(t = grid_ts, d = grid_ds)

```

```

bandwidth_df <- tibble(t_bandwidth = t_bandwidths,
                      d_bandwidth = d_bandwidths,
                      null_t_bandwidth = null_t_bandwidths)

## hyperparameters for random-grid-search Gozalo test
t_limits <- 12 * c(70, 90)
d_limits <- 12 * c(2, 4)
no_grid_points <- 15

etas <- 10^(-1:-5)
deltas <- seq(0.2, 0.8, 0.2)

hyperparameter_df <- crossing(bandwidth_df, etas, deltas) %>%
  as_tibble() %>%
  rename(eta = etas, delta = deltas)

```

## Schätzung der Intensitäten

```

library(tidyverse)

setwd("C:/Users/Radloff/Documents/Promotion/Dementia2Death/R_Code_Dementia")

source("nielsen_linton_tools.R")
source("hyperparameter_selection.R")

dementia_analysis <- read_csv("../Data/dementia_analysis.csv")

# transform data to format suited for estimation functions
# choose only data from females
female_analysis_data <- dementia_analysis %>%
  filter(!is.na(time_to_dementia), sex == "female") %>%
  mutate(
    Z1 = time_to_dementia,
    Z2 = pmin(time_to_death, time_to_censoring, na.rm = T),
    DELTA = ifelse(is.na(time_to_death), 1, 2)
  ) %>%
  select(Z1, Z2, DELTA)

female_non_dementia_data <- dementia_analysis %>%
  filter(is.na(time_to_dementia), sex == "female") %>%
  mutate(Z = pmin(time_to_death, time_to_censoring, na.rm = TRUE),
         DELTA = is.na(time_to_censoring)) %>%
  select(Z, DELTA)

# NL estimations for range of bandwidths
female_nl_analysis_fixed_bw_ratio <- female_analysis_data %>%
  grid_nl_estimate(ts = value_range(c(780, 1150), 120),
                  ds = value_range(c(0, 100), 120),
                  t_bws = t_bandwidths,
                  d_bws = d_bandwidths,
                  kernel = kernel)

write_csv(female_nl_analysis_fixed_bw_ratio,
          "../Data/final_results/female_nl_analysis_fixed_bw_ratio.csv")

# NL estimations for range of bandwidths under nullhypothesis
female_nl_analysis_under_null <- female_analysis_data %>%

```

```

grid_nl_estimate_under_null(ts = value_range(c(780, 1150), 120),
  bws = null_t_bandwidths,
  kernel = kernel)

write_csv(female_nl_analysis_under_null,
  "../Data/final_results/female_nl_analysis_under_null.csv")

# estimation of death hazards for women not suffering from dementia
female_hjort_analysis_non_demented <- female_non_dementia_data %>%
  grid_hjort_estimate(ts = value_range(c(780, 1150), 120),
    bws = hjort_t_bandwidths,
    kernel = kernel)

write_csv(female_hjort_analysis_non_demented,
  "../Data/final_results/female_hjort_analysis_non_demented.csv")

# NL estimation plus variance estimation for each point
female_nl_analysis_including_variance <- female_nl_analysis_fixed_bw_ratio %>%
  mutate(variance_est = pmap_dbl(.l = list(t, d, t_bandwidth, d_bandwidth),
    ~pointwise_nl_variance_estimate(data = female_analysis_data,
      t = ..1,
      d = ..2,
      t_bandwidth = ..3,
      d_bandwidth = ..4,
      kernel)))

write_csv(female_nl_analysis_including_variance,
  "../Data/final_results/female_nl_analysis_including_variance.csv")

```

## Durchführung der Anpassungstests

```
library(tidyverse)

source("gozalo_tests_tools.R")
source("nielsen_linton_tools.R")
source("hyperparameter_selection.R")

setwd("C:/Users/Radloff/Documents/Promotion/Dementia2Death/R_Code_Dementia")
dementia_analysis <- read_csv("../Data/dementia_analysis.csv")

female_analysis_data <- dementia_analysis %>%
  filter(!is.na(time_to_dementia), sex == "female") %>%
  mutate(
    Z1 = time_to_dementia,
    Z2 = pmin(time_to_death, time_to_censoring, na.rm = T),
    DELTA = ifelse(is.na(time_to_death), 1, 2)
  ) %>%
  select(Z1, Z2, DELTA)

## fixed-grid test

female_fixed_gozalo_test_results <- bandwidth_df %>%
  mutate(p_value = pmap_dbl(.l = list(t_bandwidth,
                                       d_bandwidth,
                                       null_t_bandwidth),
    .f = ~fixed_grid_gozalo_test(
      data = female_analysis_data,
      grid = fixed_grid,
      t_bandwidth = .l,

```

```

        d_bandwidth = ..2,
        null_t_bandwidth = ..3,
        kernel = triangular_kernel)
    )
)

write_csv(female_fixed_gozalo_test_results ,
          "../Data/final_results/female_fixed_gozalo_test_results.csv")

## random-grid search test
female_grid_search_gozalo_test_results <-
  hyperparameter_grid_grid_search_gozalo_test(female_analysis_data,
                                              hyperparameter_df,
                                              point_grid = "each_new")

write_csv(female_grid_search_gozalo_test_results ,
          "../Data/final_results/female_grid_search_gozalo_test_results.csv")

female_grid_search_gozalo_test_results_one_grid <-
  hyperparameter_grid_grid_search_gozalo_test(female_analysis_data,
                                              hyperparameter_df,
                                              point_grid = "all_same")

write_csv(female_grid_search_gozalo_test_results_one_grid ,
          "../Data/final_results/female_grid_search_gozalo_test_results_one_grid.csv")

```

## Erstellung von Grafiken

```

# This script contains code for the generation of the plots of my thesis.
# Further, a few summary statistics are computed.

library(tidyverse)
library(cowplot)
library(plotly)
library(orca)

setwd("C:/Users/Radloff/Documents/Promotion/Dementia2Death/R_Code_Dementia")

source("nielsen_linton_tools_plotting.R")
source("nielsen_linton_tools_support.R")
source("hyperparameter_selection.R")

dementia_per_person <- read_csv("../data/dementia_per_person.csv")

# compute some information for thesis text
dementia_per_person %>%
  summarize(n = n(),
            no_got_dementia = sum(!is.na(year_of_dementia)),
            perc_got_dementia = mean(!is.na(year_of_dementia)),
            perc_has_died = mean(!is.na(year_of_death)),
            perc_censored_before_study_end = mean(!is.na(year_of_censoring) &
                                                    (year_of_censoring < 2015 |
                                                     quarter_of_censoring < 4)),
            perc_censored_at_study_end = mean(!is.na(year_of_censoring) &
                                              year_of_censoring == 2015 &
                                              quarter_of_censoring == 4))

dementia_per_person %>%
  group_by(sex) %>%

```

```

summarize(n = n(),
  no_got_dementia = sum(!is.na(year_of_dementia)),
  perc_got_dementia = mean(!is.na(year_of_dementia)),
  perc_has_died = mean(!is.na(year_of_death)),
  perc_censored_before_study_end = mean(!is.na(year_of_censoring) &
                                          (year_of_censoring < 2015 |
                                           quarter_of_censoring < 4)),
  perc_censored_at_study_end = mean(!is.na(year_of_censoring) &
                                     year_of_censoring == 2015 &
                                     quarter_of_censoring == 4))

dementia_per_person %>%
  summarize(perc_female = mean(sex == "female"))

## Plot: differences between sexes in number and percentage of dementia patients
dementia_by_sex_plot <- dementia_per_person %>%
  mutate(got_dementia = !is.na(year_of_dementia)) %>%
  ggplot(aes(fill = got_dementia, x = sex)) +
    geom_bar(position = "fill") +
    ylab("proportion") +
    scale_fill_grey(start = 0.8, end = 0.2) +
    theme_bw()

distr_of_sex_plot <- dementia_per_person %>%
  ggplot(aes(x = sex)) +
    geom_bar(color = "black", alpha = 0.5) +
    theme_bw()

grid.arrange(distr_of_sex_plot, dementia_by_sex_plot, ncol = 2)

save_plot("sex_diff_plot.pdf",
  plot_grid(distr_of_sex_plot, dementia_by_sex_plot),
  ncol = 2)

## Plots of estimation and test results

# read in all the estimations and test results
female_nl_analysis <-
  read_csv("../data/final_results/female_nl_analysis_fixed_bw_ratio.csv")
female_nl_analysis_under_null <-
  read_csv("../data/final_results/female_nl_analysis_under_null.csv")
female_non_demented <-
  read_csv("../data/final_results/female_hjort_analysis_non_demented.csv")
female_fixed_gozalo_test_results <-
  read_csv("../data/final_results/female_fixed_gozalo_test_results.csv")
female_grid_search_gozalo_test_results <-
  read_csv("../data/final_results/female_grid_search_gozalo_test_results.csv")
female_grid_search_gozalo_test_results_one_grid <-
  read_csv("../Data/final_results/
  female_grid_search_gozalo_test_results_one_grid.csv")
female_nl_analysis_including_variance <-
  read_csv("../Data/final_results/female_nl_analysis_including_variance.csv")

n <- nrow(
  dementia_per_person %>%
    filter(sex == "female",
      !is.na(year_of_dementia))
)

```

```

# choice of bandwidth for subsequent plots
plot_t_bandwidth <- 24

## 3D-surface plot

alpha_plot <- female_nl_analysis %>%
  filter(t_bandwidth == plot_t_bandwidth) %>%
  select(-t_bandwidth, -d_bandwidth) %>%
  spread(key = d, value = alpha) %>%
  column_to_rownames(var = "t")

t_plot <- female_nl_analysis %>%
  filter(t_bandwidth == plot_t_bandwidth) %>%
  select(t) %>%
  distinct() %>%
  arrange() %>%
  pull()

d_plot <- female_nl_analysis %>%
  filter(t_bandwidth == plot_t_bandwidth) %>%
  select(d) %>%
  distinct() %>%
  arrange() %>%
  pull()

plot_list <- list(t = t_plot/12, d = d_plot/12, alpha = as.matrix(alpha_plot))

scene <- list(camera = list(up = list(x = 0, y = 0, z = 1),
                                eye = list(x = -1, y = -1.6, z = 0.7),
                                center = list(x = 0, y = 0, z = -0.2)))

surface_plot <- plot_ly() %>%
  add_surface(data = plot_list, x = ~d, y = ~t, z = ~alpha, colors = "Greys") %>%
  layout(scene = scene)

# doesn't work at the moment
# have saved image via export button in the viewer
# orca(surface_plot,
# "C:/Users/Radloff/Documents/Promotion/Doktorarbeit Text/Plots/surface_plot.pdf")

## two cross section plots for fixed t

fixed_ts <- 12 * c(70, 80, 90)
fixed_ts <- map_dbl(fixed_ts, ~nearest_neighbor(., female_nl_analysis$t))

null_values <- female_nl_analysis_under_null %>%
  filter(bandwidth == plot_t_bandwidth / t_to_null_t_bandwidth_ratio,
         t %in% fixed_ts)

nl_values <- female_nl_analysis %>%
  filter(t_bandwidth == plot_t_bandwidth,
         t %in% fixed_ts)

cross_sections_plot <- ggplot() +
  geom_line(data = nl_values, mapping = aes(x = d/12, y = alpha,
                                           linetype = factor(round(t/12))),
           size = 1.1) +
  geom_hline(data = null_values, mapping = aes(yintercept = alpha,
                                                linetype = factor(round(t/12))),
            size = 1.1)

```

```

        color = "grey", size = 1.05) +
    ylim(c(0, max(nl_values$alpha))) +
    labs(linetype = "t", x = "d") +
    theme_bw()

save_plot("C:/Users/Radloff/Documents/Promotion/Doktorarbeit_Text/Plots/
          cross_sections_plot.pdf",
          cross_sections_plot)

## comparison to non-demented curve
# take person who gets dementia diagnosis at age t_dementia_diagnosis

t_dementia_diagnosis <- 12 * 75
max_duration <- 12 * 5

max_age <- t_dementia_diagnosis + max_duration

hazard_without_dementia <- female_hjort_analysis_non_demented %>%
  filter(t >= t_dementia_diagnosis,
         bandwidth == 12) %>%
  select(t, alpha)

hazard_under_null <- female_nl_analysis_under_null %>%
  filter(t >= t_dementia_diagnosis,
         bandwidth == plot_t_bandwidth / t_to_null_t_bandwidth_ratio) %>%
  select(t, alpha)

nl_obs <- female_nl_analysis %>%
  filter(t >= t_dementia_diagnosis,
         t_bandwidth == plot_t_bandwidth) %>%
  group_by(t) %>%
  summarize(d = nearest_neighbor(t[1] - t_dementia_diagnosis, d))

hazard_nl <- female_nl_analysis %>%
  filter(t_bandwidth == plot_t_bandwidth) %>%
  semi_join(nl_obs) %>%
  select(t, alpha)

different_hazards <- bind_rows(list(without_dementia = hazard_without_dementia,
                                   not_duration_dependent = hazard_under_null,
                                   duration_dependent = hazard_nl),
                              .id = "type_of_hazard")

different_hazards_2 <- bind_rows(
  list(all_t = different_hazards,
       early_t = filter(different_hazards, t <= max_age)),
  .id = "plot_type"
)

indivuium_hazard_plot <- different_hazards_2 %>%
  ggplot(aes(x = t/12, y = alpha, linetype = type_of_hazard)) +
  geom_line(size = 1.1) +
  facet_wrap(~plot_type, scales = "free") +
  labs(x = "age") +
  theme_bw()

save_plot("C:/Users/Radloff/Documents/Promotion/Doktorarbeit_Text/Plots/
          individuum_hazard_plot.pdf",

```

```

indivium_hazard_plot)

## fixed grid gozalo test
fixed_test_plot <- female_fixed_gozalo_test_results %>%
  ggplot(aes(x = t_bandwidth, y = p_value)) +
    geom_line(size = 1.05) +
    geom_point(size = 1.8) +
    geom_hline(yintercept = 0.05, linetype = "dashed", size = 1.05) +
    theme_bw()

save_plot("C:/Users/Radloff/Documents/Promotion/Doktorarbeit_Text/Plots/
          fixed_test_plot.pdf",
          fixed_test_plot)

## random grid seach gozalo test

plot_etas <- c(0.00001, 0.001, 0.1)
plot_deltas <- c(0.8, 0.6, 0.2)
plot_hps <- tibble(eta = plot_etas, delta = plot_deltas)

female_grid_search_gozalo_test_results %>%
  mutate(delta = round(delta, 1)) %>%
  semi_join(plot_hps) %>%
  ggplot(aes(x = t_bandwidth, y = p_value)) +
    geom_line() +
    geom_point() +
    geom_hline(yintercept = 0.05, linetype = "dashed") +
    facet_wrap(~eta + delta)

# one grid version
random_grid_test_plot <- female_grid_search_gozalo_test_results_one_grid %>%
  mutate(delta = round(delta, 1)) %>%
  semi_join(plot_hps) %>%
  ggplot(aes(x = t_bandwidth, y = p_value)) +
    geom_line(size = 1.05) +
    geom_point(size = 1.8) +
    geom_hline(yintercept = 0.05, linetype = "dashed", size = 1.05) +
    facet_wrap(~eta + delta, scales = "free", labeller = label_both) +
    theme_bw()

save_plot("C:/Users/Radloff/Documents/Promotion/Doktorarbeit_Text/Plots/
          random_test_plot.pdf",
          random_grid_test_plot)

## Plot penalties vs S_n_sq statistics

S_n_data <- female_nl_analysis_including_variance %>%
  left_join(
    (female_nl_analysis_under_null %>%
      mutate(t_bandwidth = round(bandwidth * t_to_null_t_bandwidth_ratio),
             alpha_null = alpha) %>%
      select(t, t_bandwidth, alpha_null))
  ) %>%
  mutate(
    S_n_squared = n * t_bandwidth * d_bandwidth *
      (alpha - alpha_null)^2 / variance_est
  )

```



```

S_n_statistics <- S_n_data %>% filter(t >= t_limits[1],
                                     t <= t_limits[2],
                                     d >= d_limits[1],
                                     d <= d_limits[2]) %>%
group_by(t_bandwidth) %>%
summarize(
  median = median(S_n_squared),
  upper_pred_interval = quantile(S_n_squared, 0.95),
  lower_pred_interval = quantile(S_n_squared, 0.05)
)

eta_delta_df <- hyperparameter_df %>%
mutate(delta = round(delta, 1)) %>%
mutate(n = nrow(female_analysis_data),
       penalty = eta * (n * t_bandwidth * d_bandwidth)^delta,
       eta_delta = paste(eta, delta, sep = ", ")) %>%
semi_join(plot_hps)

penalty_plot <- ggplot() +
  geom_line(data = eta_delta_df,
            mapping = aes(x = t_bandwidth, y = penalty, linetype = eta_delta),
            size = 1.05) +
  geom_line(data = S_n_statistics,
            mapping = aes(x = t_bandwidth, y = upper_pred_interval,
                          color = "grey", size = 1.05, linetype = "dashed") +
  geom_line(data = S_n_statistics,
            mapping = aes(x = t_bandwidth, y = median,
                          color = "grey", size = 1.05) +
  geom_line(data = S_n_statistics,
            mapping = aes(x = t_bandwidth, y = lower_pred_interval,
                          color = "grey", size = 1.05, linetype = "dashed") +
  labs(linetype = "eta, Δdelta") +
  theme_bw()

save_plot("C:/Users/Radloff/Documents/Promotion/Doktorarbeit_Text/Plots/
          penalty_plot.pdf",
          penalty_plot)

## Plot showing distribution of S_n_squared across t,d for one t_bandwidth
# most likely not for thesis
t_bw <- 8
max_S_n_sq <- 1000
S_n_data %>% filter(t >= t_limits[1],
                  t <= t_limits[2],
                  d >= d_limits[1],
                  d <= d_limits[2],
                  t_bandwidth == t_bw,
                  S_n_squared < max_S_n_sq) %>%
mutate(large_S_n = S_n_squared > 1000) %>%
ggplot(aes(x = d, y = t, fill = S_n_squared)) + geom_tile()

S_n_data %>% filter(t >= t_limits[1],
                  t <= t_limits[2],
                  d >= d_limits[1],
                  d <= d_limits[2],
                  t_bandwidth == t_bw,
                  S_n_squared < max_S_n_sq) %>%
ggplot(aes(x = S_n_squared)) +
geom_histogram() +

```

```
ggtitle(paste("Bandwidth:", t_bw, "S_n_sq_Bound:", max_S_n_sq))
```

## Hilfsfunktionen

```
#' positive part of a number (a vector of numbers respectively)
#' ARGS
#'   x: vector of real numbers
#' RETURN
#'   vector of real numbers: x if x > 0, 0 else (elementwise)
pos <- function(x) {
  x * (x > 0)
}

#' range of values, equidistantly distributed across some interval,
#' without touching the limits
#' distance of the outer points from the margins
#' equal half the distances between points
#' ARGS
#'   range: vector of length 2, the first element being smaller than the second.
#'   limits of the range the values ought to be distributed across
#'   number: integer. number of values to be distributed across the interval
#' RETURN
#'   vector of length number.
#' EXAMPLE
#'   value_range(c(0,1), 2) = c(0.25, 0.75)
value_range <- function(range, number) {
  values <- seq(range[1], range[2], length.out = number + 1)
  return( ( values[1:number] + values[2:(number + 1)] ) / 2 )
}

#' vectorized triangular kernel function
#' ARGS
#'   x: numeric vector
#' RETURN
#'   numeric vector of same length as x
triangular_kernel <- function(x) {
  return(
    (abs(x) <= 1) * (1 - abs(x))
  )
}

#' vectorized uniform kernel function
#' ARGS
#'   x: numeric vector
#' RETURN
#'   numeric vector of same length as x
uniform_kernel <- function(x) {
  return((abs(x) <= 1) / 2)
}

#' vectorized "halfcircle" kernel function
#' f(x) = sqrt(1-x^2) for |x|<1 and 0 otherwise
#' ARGS
#'   x: numeric vector
#' RETURN
#'   numeric vector of same length as x
halfcircle_kernel <- function(x) {
  if_else( abs(x) > 1, 0, (1 - x^2)^(1/2) )
}
```

```
#' nearest neighbor for each number in a vector x in another given vector vec
#' Example: nearest_neighbor(c(11.1, 19.4, 13.8), 10:20) = c(11, 19, 13)
#' ARGS
#'   x: numeric vector
#'   vec: numeric vector
#' RETURN
#'   numeric vector of same length as x, containing only numbers from vec
nearest_neighbor <- function(x, vec) {
  map_dbl(x, ~ vec[which.min(abs(. - vec))])
}
```

# Anhang B

## Lebenslauf

### Allgemeines

---

Geburt      05. April 1992, Eutin.

### Bildung

---

2018–2021    Promotion am Lehrstuhl für Statistik und Ökonometrie,  
                 Universität Rostock.

2018          M.Sc. Wirtschaftsmathematik, Universität Rostock. Gesamtnote 1,1.

2016/17      Erasmussemester in Lund, Schweden.

2015          B.Sc. Mathematik, Universität Rostock. Gesamtnote 1,3.

2011          Abitur, Carl-Maria-von-Weber-Schule, Eutin.

### Arbeitserfahrung

---

2018–2021    Wissenschaftlicher Mitarbeiter  
                 Lehrstuhl für Statistik und Ökonometrie, Universität Rostock.

2017          Dreimonatiges Praktikum bei HDI Kundenservice AG, Köln.  
                 Bereich Risikomanagement

2016/17      Wissenschaftliche Hilfskraft  
                 Tutorium und Korrektur von Hausaufgaben

2011/12      Freiwilliges Soziales Jahr, Preetzer Werkstätten.